CONTEXTUAL INFLUENCES ON THE RATING PROCESS: THE EFFECTS OF ACCOUNTABILITY AND RATING OUTCOMES ON PERFORMANCE RATING QUALITY

By

NEAL P. MERO

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1994

To my parents, James and Inez Mero, who taught me the value of faith, family, and friends.

ACKNOWLEDGMENTS

With a doctoral program as well as most important things in life there are always many people who help you along the way. That was certainly the case in the completion of this work. There is no question that any success I have had is due to the patient, selfless, and unshakable love and support of my wife Cheryl. As my constant companion and friend, regardless of the path I have chosen to follow she has always been with me. I thank her for the many times when it was her confidence in me that kept me going.

I would also like to thank my two children Samantha and Brent for their patience, support, and help. Their willingness to collate and staple material, move away from the computer, and skip a school activity in support of their Dad is remarkable.

I am also blessed with a wonderful family who has provided me tremendous support throughout this program. Thanks go to my parents Jim and Inez Mero and my brothers and sisters Mary, Carole, Donna, Pam, Jim, and Mark for always being there for me. They form the core of the most "functional" family I know. I also count my dear friends Phil and Karen Doucet as part of my family and my support group.

Their friendship has been one of the true joys of my life for many years.

I owe special thanks to Professor Stephan Motowidlo. His patient intellectual nurturing, encouragement, and occasional careful redirection have contributed greatly to this project and to my intellectual development. I also am indebted to my committee members, Professors Henry Tosi, Jerald Young, Linda Crocker, and Barry Schlenker. They were always willing to listen to my ideas and point me in directions that greatly increased the usefulness of this research. I would also like to thank Professor John Hall for his constant encouragement and support of my research.

I am also grateful to my Air Force mentors Colonel James Woody, Colonel Chuck Yoos, and Lt Col Mike Wenger for their confidence and support.

There were many family members, friends, and colleagues who contributed to this research as reviewers, expert judges, or actors. I am indeed thankful for the help of Mike Wenger, Rita Campbell, Bill Paul, Tony Wolusky, Deborah Hall, Thornton Burgess, Maureen Hornyak, Marty Hornyak, Julie Boit, Phil Doucet, Karen Doucet, Kevin Banning, Allison Banning, Mary Jo Vaughn, Brian Vaughn, Jennifer Burnett, Steve Werner, Theresa Van Noy, Steve Green, Kurt Heppard, Kevin Wolfe, Jeannie Wolfe, Tammi Jackson, Ron Jackson, Cheryl Mero, Jeff Katz, Shelly Katz, and Robin Sammons. Additionally, this project could not have been completed without the encouragement and

technical support of Olen and Doris McCullough. I am also grateful to Dennis and Carrie Lee for their gracious willingness to allow me the use of the Florida Woodland facilities.

Finally, I would like to thank three special friends Steve Werner, Jeff Katz, and Kevin Banning for their encouragement, intellectual support, and friendship during this endeavor. They were always willing to help me maintain my perspective even when I didn't realize I had lost it.

TABLE OF CONTENTS

ACKNOWLEDGMENTS iii

ABSTRACT viii

CHAPTERS

1. INTRODUCTION, LITERATURE REVIEW, AND HYPOTHESES 1

Introduction 1
Literature Review 3
Hypotheses 32

2. METHODOLOGY 42

Methodological Issues for Laboratory Studies . 42
Managerial Simulation 44
Overview of Procedures Used in This Study . . 58
Subjects 60
Manipulation of Accountability and Rating Outcomes 60
Subordinate Performance Presented in this Study 62
Instruments 64
Dependent Variables 66

3. RESULTS 70

Manipulation Checks 70
Methods of Analysis 73
Tests of Hypotheses 73
Additional Analyses 84

4. DISCUSSION. 100

Findings 100
Implications for Future Research 109
Summary 112

REFERENCES 114

APPENDICES

A SAMPLE IN-BASKET ITEMS 122
B BEHAVIORALLY ANCHORED RATING SCALES 150
C EXPERT JUDGES' INSTRUCTIONS 156
D STUDY PROTOCOL 162
E ACCOUNTABILITY MANIPULATION 173
F RATING OUTCOME MANIPULATION 178
G RATING FORMAT 187
H DEMOGRAPHIC QUESTIONNAIRE 192
I MANIPULATION CHECK 194
J OBSERVATIONAL ACCURACY MEASURE 196
K POST EXPERIMENT QUESTIONNAIRE 200
L ATTENTIVENESS RATING SCALE 204
M NOTE TAKING RATING SCALE 206

BIOGRAPHICAL SKETCH 208

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

CONTEXTUAL INFLUENCES ON THE RATING PROCESS: THE EFFECTS OF ACCOUNTABILITY AND RATING OUTCOMES ON PERFORMANCE RATING QUALITY

By

Neal P. Mero

April, 1994

Chairperson: Stephan J. Motowidlo
Major Department: Management

Recent research on factors that affect the quality of performance ratings has made little progress in actually improving rater accuracy. This has led researchers to call for increased consideration of the context within which rating decisions are made. This research considered two factors of the organizational context. Rater accountability considers how raters behave when they are required to justify their ratings to others. Rating outcome considers how differences in information about the outcomes of potential ratings affect the decision process and the subsequent ratings.

Accountability, where subjects were told they would either have to justify their ratings or that their ratings would remain anonymous, was crossed with four different rating outcomes. A unique laboratory design was used that incorporated the use of in-baskets and videotaped presentations of ratee performance to increase the complexity of the rating task. Two hundred forty-seven subjects observed and rated the performance of four ratees on three performance dimension based on 52 videotaped samples of subordinate performance presented in two sessions over a two week period.

Accountable subjects were significantly more accurate than subjects who were not held accountable for their ratings regardless of rating outcome. Accountable subjects in two of the rating outcome conditions also had higher recall of specific performance events. Accountability interacted with rating outcome in two conditions as raters complied with the pressure of the situation to alter their ratings. However, these subjects' recall of performance information was not affected, suggesting that differences in rating outcome did not affect information processing even if the rating was affected. Additional structural analyses suggested that accountable raters were more accurate because they more carefully observed, recorded, and considered the implications of ratee performance. Those three factors, which mediated the effect of accountability on accuracy, accounted for over 37% of the variance in the observational accuracy and differential accuracy of subjects in a model corrected for attenuation.

Implications of the findings and suggestions for future research are discussed.

CHAPTER 1
INTRODUCTION, LITERATURE REVIEW, AND HYPOTHESES

Introduction

Despite an extensive body of research on factors that affect the accuracy of performance ratings, little progress has been made in actually improving rater accuracy (Bernardin & Beatty, 1984; DeNisi & Williams, 1988; Landy & Farr, 1980; Murphy & Cleveland, 1991). Most of the research done so far focused on the rater. It typically involved training programs for raters and new rating methods and formats designed to help improve rater's accuracy.

The limited progress achieved by this approach which focused on raters led to a call for research that would focus on the context surrounding the rating process (Ilgen & Feldman, 1983; Murphy & Cleveland, 1991). By including contextual variables in their studies of performance appraisal, researchers should be able to consider how an organization's social system influences rater information processing and decision making. Organizational contexts place many pressures on raters, pressures that can affect both the accuracy and validity of performance ratings. Partly for this reason, Murphy and Cleveland (1991) suggested that very little can be accomplished by changing the rater or the rater's task if the context influences raters to be inaccurate.

Research reported here considers two aspects of the rating context, specific rating outcomes and rater accountability. The focus on rating outcomes extends previous work on the effects of rating purpose on rater's information processing. Cleveland and Murphy (1992) argued that rating behavior is goal directed and suggested that the social context where the rating takes place influences the goals pursued by participants. As a result, performance appraisal should be viewed as a social and communication process. Previous research suggests that what raters know about the purposes or outcomes of their ratings can affect their strategies for gathering information, and the leniency, accuracy, and ultimate quality of their ratings. The possibility that various features of the context of the performance appraisal may encourage the rater to rate more leniently, more harshly, or less accurately to achieve specific purposes is a primary focus of this research.

This research also considers effects of rater accountability, a contextual variable only rarely considered in the performance appraisal literature. Accountability is the link between individual decision makers and the social system to which they belong (Tetlock, 1985). Accountability influences decision makers by causing them to behave in a way that allows them to account for their decisions

(Schlenker, 1980; 1982; Tetlock, 1985). Decision makers respond to the pressures of accountability in at least two ways. First, when they know the audience's views, decision makers make decisions they believe will be acceptable to others (Tetlock, 1985, p.311). Second, when they do not know the audience's views, decision makers will use more cognitively complex information processing strategies (Tetlock, 1985) that lead to impressions based on a more careful consideration of relevant data. Only Klimoski and Inks (1990) have considered the effects of accountability on performance evaluation.

Literature Review

This review first considers rating process research beginning with Wherry's (1952) model of the rating process, an important forerunner of subsequent performance appraisal research. Then, other models of rater cognitive processes are considered. Next, theoretical and empirical research which considered the effect of rating purpose on rating quality is discussed. Finally, research on the effects of accountability on decision makers is reviewed. This latter section focuses on how contextual variables such as accountability may lead raters to choose different information processing strategies.

Rating Process Research

Recent research attention on performance appraisal focused on cognitive processes used by raters to make rating decisions. The process emphasis stems from Landy and Farr's (1980) concern that despite research efforts into other aspects of performance appraisals, little progress had been made in understanding variables that led to accurate or inaccurate appraisals. As a result of this work, several models of rating processes were proposed (Bernardin & Beatty, 1984; DeNisi & Williams, 1988; DeNisi, Cafferty & Meglino, 1984; Feldman, 1981; Ilgen & Feldman, 1983; Landy & Farr, 1980 & 1983; Landy, Zedeck & Cleveland, 1983; Motowidlo, 1986; and Murphy & Cleveland, 1991). This literature has been criticized, however, because of its lack of practical application (Banks & Murphy, 1985) and concerns about the generalizability of laboratory studies of the rating process to field settings (Bernardin & Villanova, 1986; Locke, 1986). This review will show that despite efforts of cognitive process researchers, there is very little that can be told to practitioners that would make them more accurate raters. However, process research has improved our understanding of conditions that influence rater accuracy. This can subsequently lead to the development of more accurate appraisal systems. In this sense, the process research reviewed here is seminal to this research effort.

Wherry's model of ratings (1952)

Wherry (1952) provided a model of ratings that was the impetus of performance appraisal process research. This work was subsequently published in an edited version (Wherry & Bartlett, 1982) and proposed factors that influence performance appraisal accuracy. These propositions stem from classical measurement theory and emphasized minimizing the error component of performance ratings. To summarize, Wherry (1952) proposed that improved accuracy will come from

1. increasing ratee control over the work situation;
2. using rating scales based on ratee controlled behaviors;
3. increasing relevant rater contact with the ratee;
4. using scale items anchored by easily observed behaviors;
5. informing raters of the behaviors to be rated;
6. encouraging raters to avoid bias;
7. keeping written records of critical incidents;
8. reducing time between observation and rating;
9. increasing frequency of observations to improve recall;
10. using behaviors that are easily classified;
11. increasing the number of items on a dimension or the number and diversity of raters to reduce the random error component.

Wherry's model set the stage for much of the work that was to come 30 years later (DeNisi and Williams, 1988).

Of direct interest to this study are two proposed linkages between the rating context and rater information processing. First, rating purpose will affect rating accuracy (Wherry & Bartlett, 1982). In appraisal settings where raters know their rating will effect the ratee, they will be less accurate, whereas in settings where ratings are seen as important to the organization, raters will be more accurate (Wherry & Bartlett, 1982). Wherry (1952) predicted that ratings obtained for experimental purposes will be more accurate than ratings collected to support administrative decisions. The proposition that different rating purposes may lead to biased ratings suggests the effect rating context can have on the rating process and as a result, rating accuracy.

A second proposition suggested that having to justify ratings will also affect raters and the type of information gathered or recalled by that rater (Wherry & Bartlett, 1982). They suggested that ratings justified to the subordinate may be more lenient while ratings justified to the supervisor may lead the rater to recall performance information known to be of interest to that supervisor (Wherry & Bartlett, 1982).

Models of the rating process

Subsequent to the work of Wherry (1952), three models of rater information processes were proposed (Denisi, Cafferty & Meglino, 1984; Feldman, 1981; Ilgen & Feldman, 1983). These models focused on the steps involved in the rating process with special emphasis on rater cognitive processes. Almost every step in this process is subject to errors that could lead to inaccurate performance appraisals (Cooper, 1981). This review considers research on the rating context and on the information processing steps of information gathering, storage, recall, and judgement.

Context of the rating process. The performance appraisal context has been described as the "noisy" environment in which the cognitive processes of the rater are performed (Feldman, 1981). Three contextual factors which influence the rater's information processing have been suggested. These factors include the amount of contact rater's have with subordinates, how well they understand their job, and how much time they can devote to the process (Feldman, 1981). Ilgen and Feldman (1983) proposed that the appraisal purpose or specific outcomes linked to the appraisal could lead raters to alter their evaluations. This is consistent with Wherry's (1952) proposition that raters will evaluate differently if appraisals will be used for rewards instead of feedback.

Murphy and Cleveland (1991) also proposed rating context influences the entire rating process including the rater's judgement, rating, and evaluation. They define context as the "heterogeneous mix of factors, ranging from the social and legal system in which the organization exists to the climate and culture within the organization" (p.25).

Several researchers (Cleveland & Landy, 1983; Eder, 1989; Eder & Buckley, 1988; Murphy & Cleveland, 1991) have suggested that the interactionist perspective (Endler & Magnusson, 1976) provides a theoretical framework from which to view the interaction between the situation or context and individual judgement. This perspective suggests that behavior is a function of the feedback between the individual and the situation. Within the performance evaluation context, this perspective suggests that situational variables, as interpreted by the rater, are critical factors in determining the outcome of the rating (Murphy & Cleveland, 1989). It is perhaps not specifically the rating purpose that affects ratings but instead, the rater's beliefs about the outcomes of those ratings.

Other researchers have stressed the importance of organizational context on personnel decision making. One review of the selection interview research reported researchers had not considered how situational factors influence interviewers (Arvey & Campion, 1982). As in performance appraisal studies, interview researchers have relied on cognitive information processing theory in an attempt to understand how perceptual processes interact during an interview (Eder, 1989).

Eder (1989) encouraged researchers to consider situational factors in decision making. His theory of the influence of interview context on interviewers proposed four dimensions of context including task clarity, interview purpose, decision risk, and accountability.

Information gathering. Within the performance evaluation context raters gather the information necessary to make the eventual judgments about performance. Denisi, Cafferty, and Meglino (1984) viewed the rater as an active seeker of information. They suggested that "preconceived notions" of the rater, such as previous impressions about the ratee or information about the rating purpose, could determine the type of information sought. Rating purpose could influence both the sample of behaviors gathered and the amount of time devoted to gathering performance information. This suggests that raters may do more than just alter their evaluations as a result of the rating purpose; they may also vary how they search for information. As a result, their rating may be based on a different behavioral sample. If the rating purpose leads the rater to gather more relevant information, rating accuracy should improve. However, if the purpose leads the rater to gather less relevant behaviors or an otherwise skewed behavioral sample, then the result should be less accuracy.

This process of sampling behaviors from a true score domain was called content sampling (Cooper, 1981). Cooper felt that a large contributor to halo error could be caused by rater judgments based on a small sample of ratee behaviors. This would force raters to rely on "global impressions" of the ratee. Motowidlo (1986) proposed a sampling model of information processing and suggested that there is a true score domain which includes the population of performance information about the target. The rater, through observation or experience, becomes aware of a sample of that information. This sample may be skewed by factors that bias the rater's attention and evaluation of that information (Motowidlo, 1986).

Feldman (1981) provided a theory of how raters attend to and recognize performance information through both automatic and controlled attention processes (Sheffrin and Schneider, 1977). He proposed that raters automatically react to certain cues from the ratee (e.g. dress, sex, race, etc.) through the use of cognitive maps or well learned scripts (Abelson, 1976). As a result, behavior that is consistent with the supervisor's expectations is noted and stored automatically (Feldman, 1981). These automatic processes may lead to generalizations which may result in inaccuracy. Controlled processes occur when inconsistent information is received and becomes salient. At this point, conscious attention is required.

Murphy and Cleveland (1991) suggested that raters attend to specific behaviors as a function of three variables. First, characteristics of the behavior may draw attention to that behavior. Second, behavior which is unique to a particular context may become salient. Third, the purpose of the observation may cause the rater to seek out specific information related to achieving that purpose. Specific rating purposes may make certain behaviors salient to the rater if they provide information that is important relative to the purpose of the performance evaluation.

Two points should be highlighted about rater information gathering processes. First, context may impact ratings by affecting the type of behaviors that raters attend to. Variations in the sample of performance information can contribute to variations in the ultimate rating. Second, certain behaviors may become salient because they provide information that is important based on the rating purpose. This may trigger the use of controlled instead of automatic information gathering processes. The next section considers how differences in the types or frequency of rating errors may be due to the use of automatic or controlled processes.

Information storage and organization. The next proposed step in the rating process considers how performance information is organized and stored. Feldman (1981) proposed that individuals use categorization schemes to facilitate the storage of information gathered by automatic processes. These categories or schematas can be based on stereotypes or trait labels which carry some judgmental aspect (ie. good, bad, lazy, energetic etc.) and carry a set of related characteristics. Cantor and Mischel (1977; 1979) proposed the use of prototypes, a representative example of the category, to facilitate categories of schemata. One category may represent the rater's description of "good" workers while another may represent a composite of behaviors or cues that represents the category of "bad" workers (Murphy & Cleveland, 1991). Information that is attended to automatically is placed in a category. Since these categories may be based on stereotypes or traits, and since they vary among people, automatic processing may lead to inaccuracies on subsequent judgements.

Ilgen and Feldman (1983) suggested that many important information processing activities are done using automatic processes which are beyond the rater's conscious control. Automatic processing should result in more bias than controlled processes because of the use of faulty schematas (Lord, 1985).

Controlled processes do not necessarily imply more rationality (Bargh, 1984). They could also introduce errors into rater information processing. Some aspect of the rating situation, such as the purpose of the observation, may cause the rater to process performance information in a controlled manner. At this point, raters first attribute the cause of the behavior (Feldman, 1971). Attribution theory (Kelly, 1971; Jones, 1979; Weiner, 1972) is concerned with whether individuals attribute the behavior to internal or external factors. Research suggests that people tend to make incorrect attributions about the causes of performance. This "fundamental attribution error" (Jones & Nisbett, 1972) occurs because people tend to explain the behavior of others in terms of internal causes but explain their own behavior in terms of external causes. As a result, when raters using controlled processes attempt to determine reasons for ratee behavior they may introduce error into the rating process. Differences in attributions about ratee performance may be directly related to the overall favorability, and perhaps accuracy, of the rating.

It is also possible that when raters attribute responsibility for behavior to either the subordinate or the situation, their decision is influenced by the implications of that attribution. If a desired rating outcome is achieved by attributing ratee behavior to external causes, which protects the ratee from responsibility, the rater may tend to make that attribution. If the rater's purposes are better served by attributing the performance to the ratee, say to justify a lower rating, then that attribution may be made. This would explain how rating purpose can affect rating quality. Although the rater may not be conscious of this process (Bargh, 1984), it is possible that these controlled processes have a significant influence on the accuracy of information stored by the rater.

Feldman (1981) suggested that after raters attribute the cause of the behaviors to a source they make a second judgement about which specific category will be used to store the information (for a detailed review see Srull &

Wyer, 1979). Since these categories include judgmental value, the assignment of a behavior to a category may mean that the behavior is judged by the rater at the time of observation (Feldman, 1981).

The research reviewed above suggested that raters, using either automatic or controlled processes, make some judgement about the behavior before it is stored and that either method is subject to the introduction of errors. It was also suggested that factors such as rating purpose may influence the judgements made about ratee behavior. Errors made in the storage process should influence the quality of information recalled by the rater at the time of the rating and, as a result, affect rating quality.

Information recall and judgement. Feldman's (1981) process model described the process of seeking and recalling stored information. He suggested information is stored on a temporary "work space" during which both observed behaviors and categories are available. Over time the work space is cleared and only the categories are retained in "storage bins." When the information is needed, the categories are recalled, not the behaviors. If several categories exist, factors such as the recency of storage or the supervisor's emotional state may influence the recall process (Feldman, 1981). This categorization scheme may not only bias recall, it may also prevent subsequent information from changing the rater's judgements (Feldman, 1981).

Finally, the rater makes judgements using the recalled information (Denisi, Cafferty & Meglino, 1984; Feldman, 1981). This retrieved information is biased towards the prototypes that formed the basis for the rater's categorization schemes. Additionally, since people tend to look for information that supports a particular decision, the entire process may be biased (Feldman, 1981). The rater's search for stored information may be guided by factors such as appraisal purpose or desired outcomes. While the rater may have observed information contrary to that purpose, it may not be recalled at the time of judgement.

Because the stored information is already integrated in the form of categories, perhaps no further judgement is needed. When integration is required, recalled beliefs about the individual are "thought to be described best by a weighted - average combination of components or subsidiary evaluations also affected by salience-causing factors" (Feldman, 1981, p. 138).

Summary of the information processing model. Raters are subject to biases at each step of the rating process. However, the chance of biased ratings increases when raters must also deal with a rating context where the ultimate goal may not be to provide accurate and reliable ratings. Unfortunately, research which has focused on just the rater has ignored the significant influence of the rating context.

This research will focus on the interaction between the rater and the rating context. Two contextual factors, rating purpose outcomes and rater accountability, are considered in the next two sections.

Rating Purpose Research

Recently, Murphy and Cleveland (1991) suggested that rating behavior should be considered a goal directed process used by the rater to achieve his or her own interests. They suggested that there are wide variances in the "instrumentality" which specific performance evaluation outcomes may have for raters (see Porter & Lawler, 1968; Schwab, Olian-Gottlieb, & Henneman, 1979; Vroom, 1964). This suggestion is consistent with one study where executives reported consciously manipulating ratings to meet their specific objectives (Longenecker, Gioia, & Simms, 1987). Murphy and Cleveland (1991) also pointed out that outcomes valued by the rater may not be the same as those valued by the organization.

The suggestion that raters intentionally manipulate their ratings to achieve their own objectives is not new. Conditions within the context of the performance appraisal may even lead raters to determine the ratings that should be given at the beginning of the rating process (Longenecker, Gioia & Simms, 1987; Murphy & Cleveland, 1991). Longenecker, Gioia, and Simms (1987) suggested the typical performance appraisal environment may not lead to objective evaluation of subordinate performance. It has been argued that raters do not fail to give accurate ratings because they are incapable of accuracy but because they are unwilling to rate accurately (Banks & Murphy, 1985).

Within the performance appraisal context differences in the purpose or expected outcomes of ratings may explain variations in rating quality. Purpose has been described as "the key mechanism by which context of the rating interacts with capabilities and cognitive and judgmental processes of the rater" (Murphy & Cleveland, 1991, p.73).

Organizations use performance appraisal information for a wide variety of purposes. One study reported four commonly used purposes for performance appraisals (Murphy and Cleveland, 1991). These purposes they reported were for between-person (promotions etc.) and within person (feedback) decisions, for system maintenance (manpower planning), and for documentation (legal requirements). Each of these purposes should have different effects in terms of rater anxiety about the consequences of those ratings.

Murphy, Balzer, Kellam, and Armstrong (1984) proposed that "purpose affects ratings indirectly by affecting basic cognitive processes such as the observation, encoding, and recall of the rated person's behavior" (p.46). Citing the Landy and Farr (1980) model they argued that "raters who believe their ratings will be used to make important decisions may be more careful in attending to the behavior of the individual being rated" (Murphy et al., 1984, p.46).

Studies of purpose effects on rating quality

Several researchers have compared the results of ratings used for experimental purposes with those used for administrative purposes such as selection, assignment, or promotion. Taylor and Wherry (1951) found ratings made for administrative purposes were significantly more lenient and differentiated less between ratees than ratings provided for experimental purposes. In another study, Driscoll and Goodwin (1979) found students were more lenient when they believed that ratings would be used for administrative decisions.

In a subsequent study, Meier and Feldhusen (1979) found no effect of purpose on the leniency of student raters. They suggested, however, that their results were tentative since a simple statement of purpose may not have had the desired effect. Murphy et al. (1984) also found purpose had no effect on accuracy in evaluations of teacher performance. However, they reported their belief that rating accuracy might be affected by raters' beliefs about the rating outcomes rather than the specific purpose. They also suggested their results may be due to the lack of interpersonal consequences for the raters if they gave low ratings.

McIntyre, Smith, and Hassett (1984) considered the effects of purpose on accuracy found only small support for the proposition that purpose affects accuracy. They considered subject ratings for the same videotaped lectures presented in the Murphy et al. (1984) study but subjects were told the purpose of their evaluation was for either research, course-improvement instructions, or for hiring decisions. The authors suggested that the small effect of purpose the result of how the purpose variable was manipulated. They suggested that future research provide more explicit detail about the effects of rating decisions on the ratees.

Studies on the impact of purpose on performance ratings have yielded inconsistent results. Some researchers have suggested that the purpose manipulation may have had few personal consequences for the raters or ratees and as a result the subjects may not have adequately understood the effect of the outcomes of their decisions on subordinates. Murphy et al. (1984) suggested that it may be the rater's perception of the outcome of the ratings rather than the stated purpose that influences rating leniency. They also contended that raters attempt to avoid the "interpersonal consequences" of rating decisions and since their subjects were guaranteed anonymity, there were few interpersonal consequences (Murphy et al., 1984). These differences in the perceived consequences of the rating may account for inconsistencies in research on leniency as a function of purpose.

The above discussion is consistent with theoretical discussions of this issue. First, several researchers have posited that leniency in ratings is more likely to occur when the ratings have important consequences for the ratee (Ilgen & Feldman, 1983; Lawler, 1976; McCall & Devries, 1976; Murphy et al., 1984) or the rater (Cleveland & Murphy, 1992; Mohrman & Lawler, 1983; Murphy & Cleveland, 1991). Second, rater desire to maintain positive working relationships may be another reason for rating inflation (Longenecker, Gioia, & Sims, 1986). Rating purpose may then affect rating quality only when there are personal consequences either directly to the rater or indirectly through effects on the ratee.

Other research has supported the proposition that rating distortion occurs when raters are required to share their ratings with subordinates (DeNisi, Cafferty & Meglino, 1984; Klimoski & Inks, 1990; Longenecker, Gioia & Sims, 1986; Waldman & Thornton, 1988). In one study, ratings used for administrative purposes were more lenient than ratings used for research purposes. Within this study a third group was told their ratings would be used for research purposes but they were also told that their ratings would be seen by subordinates. Ratings for this group were as lenient as ratings provided solely for administrative purposes (Waldman & Thornton, 1988).

Empirical evidence suggests that the effects of rating purpose on performance ratings may be moderated by whether or not there are personal implications for the rater or the ratee. Rather than providing raters vague purposes for their ratings, this study will provide raters specific information about how rating decisions affect both the organization and the ratee. This is similar to actual appraisal situations where raters see how their rating decisions directly impact their subordinates, their unit, or perhaps more importantly themselves. The contextual cues provided by information about the outcomes of ratings may simply lead raters to provide ratings that will achieve preferred outcomes.

Effects of rating purpose on rater cognitive processes

There is empirical support for the possibility that the effect of purpose on performance is through its influence on rater cognitive processes. One study reported that raters searched for different types of information under different conditions of appraisal purpose. Raters required to make specific decisions such as giving an employee a raise and determining the amount of that raise, searched for the normative data necessary to support such a decision (Williams, DeNisi, Blencoe & Cafferty, 1985). Williams et al. (1985) suggested rating purpose or outcome may serve to do more than just motivate raters to provide a desirable rating. It may also influence rater cognitive processing of performance information. Eder's (1989) proposed effects of context on interviewers also suggested that purpose may influence the type of information that is sought and stored. Raters may determine in advance what kind of information is needed and then they search for that information (Crocker, 1981).

There have been several investigations into the proposition that rating purpose influences information processing. Using a policy capturing technique, Zedeck and Cascio (1982) found that raters combined information differently when ratings were to be used for different purposes. Here, raters were given one of three purposes, merit raise, developmental, or retention. Each purpose accounted for differences in how raters weighed, combined, and integrated performance dimensions. A study on the effects of purpose on rating accuracy (Murphy, Philbin, & Adams, 1989) found that subjects told to evaluate a lecturer's performance evaluated that performance more accurately than subjects told to rate the lecture content. However, when time between performance observation and evaluation was delayed (7 days), the evaluation group's ratings were poorer than those of the content group. Murphy et al. attributed this to the difference between on-line and memory based information processing and suggested that rating purpose affects the encoding and retrieval of information about ratee behavior.

A related line of research considered how raters process information when performance appraisal is not their main concern. Balzer (1986) found that when the appraisal task was salient, it interacted with the rater's impression of ratees to influence how information was attended to and encoded. In two other studies (Hastie & Park, 1986; Williams, 1984), researchers found that when appraising individual performance is a salient purpose of observation, raters used person categories to organize information.

Summary of purpose literature

A review of the research on the effects of rating purpose suggests several observations. First, while the theoretical basis for purpose affecting ratings and rating processes has been established, the empirical research supporting that linkage is equivocal. It seems that purpose may affect ratings because it causes the rater to be concerned about specific rating outcomes. In the research cited above, subjects were typically given generic purposes (i.e. research, administrative decisions, etc.) rather than exposed to more specific outcomes of their ratings. As a result, it was up to the subject to determine specific outcomes of their decisions.

Second, researchers suggest that rating purpose may affect rating quality when there are personal implications for the rater. Subjects were usually told their ratings would remain anonymous, a condition that is unlikely to occur in organizations where both the rater's supervisors and subordinates generally have access to performance ratings. One way that raters are personally linked to their ratings is when they have to justify those decisions to others.

Third, researchers should consider whether contextual factors influence rating quality due to its affects on rater cognitive processes. Just because the ultimate rating is affected by rating purpose does not mean that raters are processing information differently. Perhaps raters process information consistently regardless of rating purpose or outcome. In this case, the effect of purpose may only be on which rating is assigned. More research is needed to determine the relationship between rating outcomes, rater information processing, and the ultimate rating decision.

Accountability Research

Wherry (1952) proposed that requiring raters to justify their ratings will affect the ratings given and the type of information gathered or recalled. This proposition has received some theoretical but little empirical support in the performance appraisal literature. However, in recent years an extensive body of literature has considered this proposition in other decision making environments.

Researchers have considered the effects of accountability on the information processing strategy of subjects (Tetlock, 1985). Results of this literature provide different pictures of the effects of accountability on decision makers. When decision makers know in advance the type of decision they can best justify, they will reach that decision. When there are no clues about which decision is easiest to justify, decision makers are thought to use a more cognitively complex decision process. While the information gathering process may be more thorough, decision makers under this condition may also consider irrelevant information. The next two sections review this research.

Decision makers as cognitive misers

When their audience's views are known, decision makers make decisions consistent with those views (Klimoski & Inks, 1990; Tetlock, 1985). Tetlock (1985) contended accountable decision makers who are aware of the views of their audience employ an "acceptability heuristic". This is based on the assumption that people are basically lazy, "cognitive misers", who prefer solutions that involve the least effort (Taylor & Fiske, 1978; Tetlock, 1985). This view suggests that accountable decision makers who know the views of the audience will make decisions they feel will be acceptable (Tetlock, 1985). As a result, these decisions are based on how justifiable they are to the audience.

There is considerable empirical support for this position. Several studies have supported the influence of the need for approval (Jones & Wortman, 1977; Wortman and Linsenmeier, 1977) and the motivation of individuals to present themselves as positively as possible to those to whom they are accountable (Baumeister, 1982; Schlenker, 1980). Tetlock (1983) showed that accountable subjects who knew the views of their audience shifted their opinions to be consistent with those views. Similar results were found in another study where subjects who were aware of views of their audience relied on a low effort acceptability heuristic and again shifted their views towards those of the audience (Tetlock, Skitka and Boettger, 1989).

Decision makers using complex decision strategies

The acceptability heuristic suggests a second view of information processing which occurs when decision makers are unaware of their audience's views. In this condition they are thought to use more complex information processing strategies. Tetlock proposed three effects on decision makers held accountable to audiences with unknown views. First, they will utilize more cognitively complex decision making strategies. Second, they will be more sure of their own cognitive processes. Third, decisions will be based on a more data driven process of impression forming (Tetlock, 1985).

Schlenker (1986) expanded on the effect of accountability on information processing. He suggested accountability leads to a) a more thorough search for relevant information, b) greater learning and recall of information, c) more complex judgement and decision strategies, d) more data-driven processing of information in which judgements are influenced by details of the information, and e) a greater awareness of decision making strategies. He also proposed the effects of accountability are attenuated when tasks are unchallenging or tedious and when the goal is merely to convince a superior that the task is completed rather than some more important goal. Consistent with ideas considered within the performance appraisal literature, the effect of accountability will be attenuated if the task has few personal implications (Schlenker, 1986).

There is empirical support for the use of more complex decision strategies. Tetlock (1983) reported that decision makers, unaware of the views of their audience, used "pre-emptive self-criticism" by developing counter-arguments to potential critics of their decisions. Simonson and Nye (1992) found that although accountable subjects used a more multi-dimensional and self-critical information processing strategy, they still selected the response expected to be more positively evaluated by the audience.

Ashton (1992) found that auditors required to justify their rating were more accurate and consistent on a task of predicting bond ratings. Telock and Kim (1987) found that accountable decision makers formed more integratively complex impressions and made more accurate predictions.

Accountable decision makers do not always make more accurate decisions. Tetlock and Boettger (1989) found that in some environments, accountability exacerbated judgmental biases through a "dilution effect". Accountable subjects used a wide range of information, including irrelevant information, when making judgements. A second study found similar results. Gordon, Rozelle, and Baxter (1988) reported that accountable interviewers considered less relevant information (in this case, the applicant's age) when making hiring decisions.

In contexts where only relevant information is presented, increased attentiveness by accountable decision makers should improve decision quality. In environments where decision makers must distinguish relevant from irrelevant information, accountability may result in biased decisions. Within most organizational decision making contexts, including performance appraisal, decision makers are inundated with information, much of which has no bearing on the decision. As a result, the possibility that a dilution effect may influence the quality of ratings provided by accountable raters must be considered.

Accountability and rater motivation

The accountability research suggests an interesting and important relationship between accountability and the subsequent motivation of the decision maker. It supports the theory that accountable decision makers adopt a decision strategy that facilitates finding an acceptable solution with a minimum of effort. This view of accountability as a motivator of certain information processing behaviors is consistent with several theories on human motivation.

One theoretical connection between the behavioral effects of accountability and motivation theory is through the evaluative nature of accountability. The requirement to account for their decisions leads decision makers to consider the criteria which they must use when they justify their decision. This evaluative aspect of accountability may, other things being equal, motivate individuals to exert greater effort (Shepperd, 1993). This is consistent with social loafing research which has shown that when individuals feel their inputs are identifiable, they are less likely to loaf (Williams, Harkins & Latane, 1981). When inputs are not identifiable, it has been proposed that motivational pressure is diminished since individuals are not held accountable for their efforts or results (Latane, Williams & Harkins, 1979; Shepperd, 1993).

A second way in which accountability is related to theories of motivation is through the explicit or implicit rewards that exist in most organizational contexts. For example, suppose raters had to account to their supervisors for the quality of their performance ratings. Those raters would be well aware that supervisors often hold considerable power in the form of organizational rewards. In situations where rating decisions must be justified to audiences other than supervisors, raters may feel anxious about the effect negative ratings would have on social relationships within the organization (Klimoski & Inks, 1990). The relationship between organizational rewards (either social or economic) and performance has been established elsewhere (Mitchell, 1974; Vroom, 1964). At the core of this relationship is that when rewards are associated with performance there is an underlying sense of evaluation that exists prior to achieving those rewards. Schlenker (1986) argued that goal setting is related to accountability because accountability suggests evaluation and goals provide a clear standard of measurement.

Questions considered in this research are developed from this relationship between accountability, motivation, and evaluation. Accountability, the within context of performance appraisal, can be viewed as motivating force on the rater. Other research suggests that raters held accountable for their ratings will consider the views of those to whom they are accountable when choosing a strategy for approaching the rating process. Lacking specific direction, raters may be sensitive to other cues from the organizational context when they determine acceptable rating behavior. Those contextual cues may be in the form of specific rating outcomes that are desired by the audience of the ratings.

Summary of accountability literature

In summary, there is considerable support for the idea that accountability affects the way decision makers process information. If the audience's views are known, decision makers make decisions consistent with those views. In doing so, they may rely on heuristics or stereotypes to communicate more easily with the audience. When the views of the audience are unknown, the decision makers use a more cognitively complex information processing strategy. It has been shown that this strategy improves the accuracy and consistency of decisions, but may also lead to decision biases when decision makers consider irrelevant information. The effects of accountability are consistent with motivation research that has demonstrated a strong and consistent relationship between the sense of evaluation and subsequent behavior.

While Klimoski and Inks (1990) are the only researchers to consider the effects of accountability in performance evaluations, Bernardin forcefully underscored the importance of accountability when he wrote "Perhaps this single factor, holding raters accountable for their ratings, just as they are held accountable for other expensive organizational resources, will do more to improve the effectiveness of a PA system than any other technique or intervention that could be recommended" (Bernardin, 1986, p.302).

Hypotheses

Although previous research has considered how generic statements of rating purpose affect rating quality, one objective of this research is to consider the effect of more specific rating outcomes on measures of rating quality. The term rating outcome as used in this study refers to a specific result of a rating. Rating outcomes are the specific implications the rating has for organizational constituents. For example, if promotion decisions are based on performance ratings, the outcome of a favorable rating would be a promotion for the ratee. The outcome of an unfavorable rating would be that the ratee is not promoted. In this example, a desirable rating outcome for the rater may be to see his or her subordinates promoted. To achieve this outcome the rater may feel pressure to provide a highly favorable rating. This may lead the rater to search for only positive examples of performance. In another situation raters may be placed in a context where a desirable rating outcome is to provide performance ratings that are useful for organizational decision making. This outcome can be achieved by accurate ratings. In this case raters should attend to a wider sample of behavior and carefully consider the impact of each behavior on the ultimate rating. Likewise, if a desirable rating outcome can be achieved by providing inflated ratings, raters may give greater weight to positive performance information.

The following predictions summarize the expected relationship between the key independent and dependent variables used in this study.

> 1. Accountable raters will provide more accurate performance ratings and remember behavioral episodes better than non-accountable raters except in conditions where desirable rating outcomes can be achieved by providing inaccurate ratings.
>
> 2. Different rating outcomes will only affect the quality of performance ratings and rater information processing when raters are personally linked to their performance ratings by the accountability manipulation.

Before presenting specific hypotheses it is appropriate to provide an overview of the manipulations used in this study. Detailed information on study design is provided in Chapter 2. In this study the rating context was manipulated by creating two accountability conditions, accountable or not accountable, crossed with four rating outcome conditions. The rating outcome manipulation provided raters information about negative outcomes that had occurred because of previous rating decisions. Raters were then encouraged to make sure their ratings led to more desirable outcomes. For example, in one condition subjects were told that the previous supervisor's tendency to provide low ratings has led to an outcome where the unit's subordinates were not competitive for promotion. Subjects were made aware of this outcome and were told that a desirable rating outcome was for their subordinates to be more competitive for promotions. In a second condition subjects were told that an outcome of previous performance ratings was that female subordinates were rated lower than male subordinates. In this condition subjects were told that a desirable rating outcome was for female subordinates to receive ratings equivalent to their male colleagues. In a third condition subjects were shown that past performance ratings had been inflated to the extent that almost all subordinates received the highest ratings. They were told that the outcome of that situation was that the organization could not use performance ratings to make important personnel decisions. In this condition subjects were told that a desirable outcome was to have performance ratings that accurately identified individual strengths and weaknesses. Subjects in the fourth outcome condition, which was used for comparison purposes, differed on the accountability manipulation but did not receive information about rating outcomes.

Rating Accuracy

The first measure of rating quality considered is rating accuracy. While several different accuracy measures have been suggested (Sulsky & Balzer, 1988), most researchers have typically used one or all of the four accuracy measures introduced by Cronbach (1955). Several researchers have pointed out that because these accuracy measures are not correlated, different measures can lead to different research conclusions (Murphy & Balzer, 1981; Murphy & Cleveland, 1991; Sulsky & Balzer, 1988). As a result, the conclusions suggested by a study of accuracy may depend more on researcher's choice of an accuracy measure than on the effects of the manipulations being studied (Becker & Cardy, 1986; Murphy & Cleveland, 1991). In this study, Cornbach's (1955) measure of differential accuracy was used. Differential accuracy measures the difference between subject ratings and true score ratings of the strengths and weakness of each ratee on each performance dimension. This measure is appropriate because the focus of this study is to determine the effect of accountability on the subject's ability to process complex performance information. The most rigorous test of this ability to make detailed distinctions in performance variations is measured by differential accuracy.

If accountability improves rater information processing, then it should be evident in the complex task of evaluating individual performance. Improved information processing by accountable raters should lead to more accurate performance ratings. There are three reasons why this should occur. First, because subjects in this study were provided several cognitively complex tasks, the justification requirement will make the performance appraisal task salient. Second, subjects who have to justify their decisions should process performance information in a way that focuses their attention on the most relevant information (Tetlock, 1985). Third, within the setting of performance appraisal, accountability should improve rating judgement by increasing the consistency with which raters process multiple pieces of performance information (Ashton, 1992).

HYPOTHESIS 1: Accountable raters will rate more accurately than non-accountable raters except when placed in a rating context where desirable rating outcomes can only be achieved by providing inaccurate ratings.

As suggested in hypothesis 1, accountable raters in some conditions may not rate more accurately. If the rating context suggests that specific ratings will lead to a desirable rating outcome, accountable raters should provide those ratings since those are easiest to justify. To achieve this desirable outcome raters should attend to, store, and recall performance information in way that makes it easier to justify the desired ratings. In other words, accountability should motivate raters to process performance information in a way that supports ratings consistent with desired outcomes.

Accountability may serve to do more than enhance rater motivation, accountability may interact with the desired outcome to cause the rater to choose a decision strategy that allows them to adequately account for their decisions. This is consistent with research that suggests that rating outcomes will only affect raters when there are personal implications--such as when they are held accountable. As a result, when accountable raters are placed in a situation where a desirable outcome can be achieved by providing ratings that disregard actual differences in subordinate performance they should provide inaccurate ratings. Raters who are not held accountable for their ratings should not feel pressured to provide ratings that achieve desirable rating outcomes. In this study this situation is created when accountable raters are encouraged to provide equivalent ratings for both male and female subordinates.

HYPOTHESIS 2: Accountable raters placed in a rating context where a desirable rating outcome can be achieved by rating female subordinates higher than their true level of performance warrants will be less accurate than non-accountable raters in the comparison group.

One group of subjects in this study was placed in a rating context where desirable rating outcomes can be achieved only if raters make accurate ratings. The interaction between the pressures of accountability and the desire to achieve outcomes facilitated by accurate ratings should result in improved rating accuracy.

HYPOTHESIS 3: Accountable raters placed in a rating context where a desirable rating outcome can be achieved by rating accurately will be more accurate than raters in the comparison group regardless of accountability status.

Within the current study, subjects in another condition were given a desired outcome that could be achieved by artificially inflating performance ratings. While this manipulation should affect rating inflation, it should not affect rating accuracy. This is based on previous research which has shown that rating inflation is not correlated with rating accuracy (See Murphy & Cleveland, 1991).

Information Processing and Recall Accuracy

This study also considered whether raters in various rating contexts processed performance information differently as indicated by how well subjects recalled specific instances of observed performance. Focusing on how accurately subjects recall performance information will provide insight into possible differences in raters information processing.

Hypotheses 1 and 3 above predicts that accountable raters in certain conditions will be more accurate raters. This result should occur because accountable subjects more carefully attend to behavioral episodes and consider how they will use information from those episodes to justify their performance ratings. If accountability leads to these improved information processing strategies, then accountable subjects in these conditions should more accurately recall specific performance information.

HYPOTHESIS 4: Accountable raters will more accurately recall performance episodes than non-accountable raters except when placed in a rating context where a desirable outcome can be achieved by providing inaccurate ratings.

Previous research has found some effects of rating purpose on rating quality. What is not understood is why this effect occurs. Perhaps when raters face a situation where accurate ratings would lead to an undesirable outcome they respond to the pressure by giving an inaccurate rating that they know does not reflect true subordinate performance. Another possibility, consistent with research cited above, is that raters selectively attend to information that facilitates the desired outcome. In this case, raters may be unaware that the ratings they are assigning are based on an biased sample of performance information. For example, if raters are in a context where they want to provide a positive rating, automatic information processing may lead them to attend to only positive information about subordinate performance. If inconsistent information is presented (such as negative performance information) raters may react using controlled processes and may choose to attribute the poor performance to situational factors beyond the control of the ratee. In either case, when raters recall information to facilitate decision making, the judgement and rating may be based on a flawed sample of behaviors that have been distorted to meet the desired purpose of the rater.

In this study, subjects in one condition were placed in a situation where a desired outcome could be achieved by rating female subordinates higher than their true performance level. If, as a result of the outcome manipulation, subjects in this condition search for positive performance information and attribute any negative performance information to external causes, this should affect their ability to accurately recall performance information. Therefore, hypothesis 5 predicts

HYPOTHESIS 5: Accountable raters placed in a rating context where a desirable outcome can be achieved by rating female subordinates higher than their true level of performance warrants will have poorer recall of performance episodes than non-accountable raters in the comparison group.

Rating Favorability

Some subjects were placed in a context where a desirable outcome could be achieved by inflating the performance ratings of all subordinates. In this case there should again be an interaction between the inflationary outcome and accountability conditions. Raters who are not required to justify their ratings should not be affected because they have no personal involvement in the rating outcome. However, accountable raters who have to justify their ratings will be more influenced by different outcomes.

HYPOTHESIS 6: Accountable raters who are placed in a rating context where a desirable rating outcome can be achieved by assigning inflated ratings will rate subordinates more favorably than non-accountable raters in the comparison group.

When subjects are presented with a situation where a desirable rating outcome can be achieved by artificially inflating the ratings of female subordinates, rating inflation is again of interest.

HYPOTHESIS 7: Accountable raters who are placed in a context where a desirable rating outcome can be achieved by assigning inflated ratings to female subordinates will rate females more favorably than non-accountable raters in the comparison group.

Dilution Effect

A final research question considers the possibility of a dilution effect as reported by Tetlock & Boettger (1989). This is the possibility that accountable raters will use irrelevant information when making rating decisions. No specific hypothesis is made concerning this dilution effect. However, if accountable subjects give weight to irrelevant information, they should be less accurate. To further test for this effect subjects were asked to recall the number of positive and negative scenes they remembered seeing for each subordinate. From this measure, two variables were created for each subject based on total number of positive and negative scenes recalled. Group means on these variables were for additional analysis of the dilution effect.

CHAPTER 2
METHODOLOGY

This chapter details the methods and procedures used for this study. It begins with a review of the methodological criticisms that have been directed at previous laboratory investigations of performance appraisal processes and effects. In response to these criticisms, this research incorporates a unique "technology" designed to allow for careful consideration of the observation and evaluation of individual performance within a controlled laboratory setting. The development of this technology is also discussed in this chapter. Next, information about the study's participants, as well as details about the study's design and procedures, are presented.

Methodological Issues For Laboratory Studies

The methods used in this study address several criticisms of previous laboratory studies which investigated the performance appraisal process. Researchers criticize laboratory studies of performance appraisal because those studies failed to consider the complexity of actual performance evaluation environments (Ebbesen & Konecni, 1980; Funder, 1987; Ilgen & Favero, 1985). Funder (1987) suggested that "research must let subjects judge real people in real social contexts, and use realistic external criteria for determining when the judgements are right or wrong" (p. 76). Another criticism by Ebbesen and Konecni (1980) argued that people may not make judgements in the real world the same way they do in the lab. Many of these criticisms were directed at earlier studies that created "paper people" or provided limited presentations of relevant ratee performance. Ilgen and Favero (1985) criticized this paper people paradigm and suggested that procedures which incorporated actual observation of performance directly or through videotaped vignettes was a promising method.

The methods used in this study address the above concerns in several ways. First, subjects were exposed to performance information on their subordinates during two 2-hour sessions over a 2 week period. This is significantly more involved than previous studies, and the extended period of time better reflects an actual rating environment. Second, participants were assigned a series of complicated administrative tasks which placed raters in a more realistic supervisory position where rating subordinates is one of many management tasks that require their attention. Third, subordinate behavior was depicted on videotape which provided raters the opportunity to observe subordinates in a wide range of diverse contexts common in modern organizations. These behavioral episodes included both rating relevant and irrelevant behavior. This technique also provided the opportunity to introduce several different forms of feedback (i.e. direct observation, viewing written work samples, observations from others etc.) on actual ratee performance. The result is a rating context that reflects an actual work environment.

Critical to any study of accuracy is the need to establish a "true score" of ratee performance. The design of this study, while limited by the constraints of a laboratory environment, had the advantage of being able to present carefully controlled information on subordinate behavior (for forming true scores). This is necessary to establish a measure from which to evaluate rating accuracy.

Managerial Simulation

This study used a managerial simulation developed to support laboratory investigations of performance appraisal settings. The simulation consists of in-basket exercises to simulate the administrative aspects of a manger's job (see Frederiksen, Jensen, & Beaton, 1972) and a series of videotaped performance episodes for subordinates introduced within the in-basket.

In-basket Description

The in-baskets place subjects into the role of Leslie Wilder a divisional manager in a federal purchasing agency.

In this position subjects managed the activities of 5 subordinate managers, Alice Garcia, Bill Jensen, Carole Donaldson, David Fredericks, and Ed Montage. These subordinate managers are responsible for the division's 5 sections and each section has a staff of 3 to 5 employees.

The in-baskets included administrative items designed to provide challenging tasks which required considerable cognitive effort and to acquaint subjects with their management responsibilities within the organization and with their subordinates. Administrative items in the in-basket were representative of the type of paperwork that typically confronts mangers on a day-to-day basis. Items such as memos from their supervisor, fellow division chiefs, subordinates, or customers were a significant part of the in-basket. These memos would ask the raters to resolve or provide input on organizational problems, develop policies, participate in special projects, handle human resource issues, schedule activities etc. Subjects were told to take all actions they thought were necessary to resolve a problem. These actions might include drafting a letter to an unhappy customer, keeping subordinates abreast of important information, reviewing trip reports, developing plans to resolve personnel problems, or advising their supervisor on important policy matters. While in-basket information was used to engage the subjects in a complex task, care was taken to insure that the in-basket presented only generic information about each of the subordinates whose behavior was to be rated. Copies of the in-baskets are provided in Appendix A.

Videotape Description

While subjects were working on the in-basket, they were shown several short videotaped vignettes of activities occurring within the organization. Because the in-basket scenario typically had subjects working at their desks, the videotapes would show someone coming in to their office to present information related to problems presented in the in-basket. Each of these vignettes was developed to present performance information about one of the subordinate managers. For example, one scene showed Alice missing a deadline by turning in a written report late. Another scene showed David dealing positively with one of his subordinates.

Videotape Development

A total of 52 separate scenes were used to present the total sample of performance information for all of the subordinate managers. Each scene was performed from a script written to reflect on the performance of one subordinate on one of the four performance dimensions of administration, coordination and negotiation, supervision, and work effort.

To guide development of the scripts, a behaviorally anchored rating scale was developed for each performance dimension. Each dimension was scaled from 1 to 7 with 7 representing the most effective performance. High, medium, and low points on the scale have corresponding behavioral anchors. This scale served two purposes. First, it provided behavioral anchors to describe examples of effective and ineffective subordinate performance that were related to the values on the rating scale. Second, these behaviors formed the basis for the behaviors that would be depicted in the vignettes. Rating scales used for scene development were also used by subjects to provide performance ratings. These scales are shown in Appendix B.

The performance scenes used in this study were chosen from a larger sample of 128 scenes developed to provide a comprehensive sample of subordinate performance that could be drawn from to support studies like this one. As stated previously, each scene demonstrated effective or ineffective performance for one subordinate on one performance dimension. While the in-basket presented 5 subordinates, situations were developed to reflect on only 4 of them. The fifth subordinate, Ed, was used to facilitate personnel problems presented in the in-baskets. Performance for each subordinate was presented in 16 situations with 4 different situations being used to reflect on each of the 4 performance dimensions. The 16 situations per subordinate resulted in the preparation of 64 situations. Each situation placed the "target subordinate" in a situation where his or her actions could realistically be viewed by their supervisor, which was to be the study participants. For each of the 64 situations, two scripts were written, one showed the subordinate performing effectively in the situation and the other showed ineffective performance. In all, a total of 128 scripts were prepared.

Scripts were then given to 4 graduate students who were asked to read each script and then identify which subordinate(s) in the scene was behaving in a way that reflected on his or her job performance. They were also asked to identify which performance dimension(s) was reflected on by the subordinate behavior. For realism purposes several scenes showed more than one subordinate even though the intent of each scene was to reflect on one subordinate's performance on one dimension. The results of this review were used to modify scripts to insure clarity of both target subordinate and dimension.

Permission was obtained from a local business to use their facilities to film each scene. Actors were chosen from available volunteers and consisted of fellow graduate students and a faculty member, as well as family and friends of this researcher. In all over 20 actors and actresses were involved.

True Score Development

Because the intent of each scene was to communicate either effective or ineffective subordinate performance, the behaviors depicted in the scene were chosen from the list of effective or ineffective behavioral anchors associated with the appropriate performance dimension. As a result, the intended true score of each scene was either a 6 if it depicted effective behavior or 2 for ineffective behavior. Since each video taped scene was to be used as a component of each subordinate's overall performance level, two studies were conducted to evaluate whether the intended value of each scene could be determined by independent samples.

Study A

Study purpose. This study's purpose was to use the experience of independent expert judges to provide feedback about the subordinate performance depicted in each of the videotaped scenes. There were two objectives. The first objective was to insure that in each scene only performance information on the one intended target subordinate on one target dimension was depicted. A second objective was to get expert rater assessment of the scale value of the behavior, based on the BARS scale prepared for this effort.

Participants. Twenty-two judges were used. Each judge had prior management experience and most judges had previous rating experience. Judges were chosen from a variety of backgrounds including academic, government, and business.

Procedures. Completed videotaped scenes were divided into four sets and each judge was asked to evaluate 32 of the 128 scenes. Videotapes were distributed to 20 expert judges who were asked to rate each subordinate in each scene whose performance reflected on one of the four performance dimensions. Judges were allowed to observe the video-tape as many times as they desired and was asked to treat and rate each scene as an independent source of performance information. Each scene was observed by 5 expert judges. A copy of the instructions sent to the expert judges is provided in Appendix C.

Results. The first consideration when reviewing Study A's results was to determine whether judge's identified the intended target subordinate and the target dimension in each scene. In 108 of 128 scenes all expert judges identified and rated the intended target dimension and subordinate. Of the remaining 20 scenes, all but one of the judges rated the intended target subordinate and dimension.

The next step was to see if the judge's ratings of the performance information in each scene were consistent with the intended rating value of a 6 for effective scenes and a 2 for ineffective scenes. Table 2-1 provides the results of the judge's ratings for all scenes. Each scene was rated by 5 judges and the mean rating and standard deviation of the ratings are shown. For each of the 64 situations there is an A scene which presented the target subordinate performing effectively on the target dimension and a B scene which presented ineffective performance.

TABLE 2-1
EXPERT JUDGE'S MEAN RATINGS AND STANDARD DEVIATIONS

SCENE	RTG	SD	SCENE	RTG	SD	SCENE	RTG	SD	SCENE	RTG	SD
1a	5.6	.89	17a	6.0	.71	33a	5.6	1.67	49a	6.0	.00
1b	1.6	.55	17b	1.6	.55	33b	1.8	.45	49b	2.0	.00
2a	6.4	.55	18a	6.2	.45	34a	6.0	.82	50a	6.0	.00
2b	2.2	.45	18b	1.6	.89	34b	1.5	.58	50b	1.5	.71
3a	6.4	.55	19a	6.4	.55	35a	4.0	.71	51a	6.4	.55
3b	2.2	1.11	19b	1.6	.55	35b	1.6	.55	51b	1.4	.55
4a	6.2	.84	20a	5.6	.55	36a	6.6	.55	52a	6.4	.55
4b	2.4	.89	20b	1.4	.55	36b	1.2	.45	52b	1.4	.55
5a	6.8	.45	21a	6.4	.55	37a	6.2	.45	53a	6.2	.45
5b	1.4	.55	21b	1.4	.55	37b	1.6	.89	53b	2.0	.71
6a	6.6	.55	22a	6.6	.55	38a	6.0	.71	54a	6.3	.50
6b	1.6	.55	22b	1.4	.55	38b	2.0	.00	54b	1.8	.50
7a	6.4	.55	23a	5.4	.89	39a	5.8	.84	55a	6.0	.71
7b	1.8	.45	23b	1.6	.55	39b	1.0	.00	55b	1.6	.55
8a	6.8	.45	24a	6.6	.55	40a	7.0	.00	56a	6.0	.82
8b	1.2	.45	24b	1.8	.45	40b	2.0	.00	56b	2.0	.00
9a	6.4	.55	25a	6.3	.50	41a	5.8	.84	57a	6.2	.45
9b	1.4	.55	25b	1.4	.55	41b	1.8	.45	57b	1.4	.55
10a	6.8	.45	26a	6.2	.45	42a	6.0	.71	58a	6.6	.55
10b	2.0	1.23	26b	1.0	.00	42b	1.4	.55	58b	1.6	.55
11a	6.6	.89	27a	5.5	.45	43a	6.2	.45	59a	6.4	.55
11b	3.6	1.95	27b	1.5	.58	43b	1.6	.55	59b	2.0	.00
12a	6.6	.55	28a	6.2	.84	44a	6.6	.55	60a	5.6	.55
12b	1.4	.55	28b	1.8	.45	44b	2.2	.45	60b	3.0	1.41
13a	6.2	.84	29a	5.6	.89	45a	6.3	.50	61a	6.2	.45
13b	1.3	.50	29b	1.6	.55	45b	1.8	.50	61b	1.6	.55
14a	6.8	.45	30a	5.6	.55	46a	5.6	1.14	62a	6.2	.45
14b	1.4	.55	30b	2.3	1.89	46b	1.5	.58	62b	1.4	.55
15a	6.0	.82	31a	6.4	.55	47a	6.8	.45	63a	6.3	.50
15b	1.6	.55	31b	1.2	.45	47b	1.4	.55	63b	1.4	.89
16a	6.0	.71	32a	5.8	.45	48a	6.4	.89	64a	6.4	.55
16b	1.8	.50	32b	1.8	.84	48b	1.8	.50	64b	2.00	.00

Mean scores for judge's ratings for high performance scenes (A scenes) was 6.18, with a mean standard deviation of .60 and a range of from a low of 4 to a high of 7. Mean score for judges' ratings for low performance scenes (B scenes) was 1.69, mean standard deviation was .57, and the range was from a low of 1 to a high of 4.

Conclusions. The results of this study suggest that each scene depicted performance information about the intended subordinate on the intended performance dimension. These results also support the contention that each scene presents the intended level of effective or ineffective performance information about the simulation's subordinates.

Study B

Study purpose. One purpose of this study was to determine whether two versions of the entire simulation, which consisted of a different set of the 64 videotaped scenes, manipulated the performance level of the subordinates as intended. A second purpose was to statistically consider alternative definitions of a true score for subordinate performance presented in the simulation.

Participants. Eighty-seven MBA students enrolled in a course on organizational behavior participated in Study B. All participants were volunteers and received course credit for their participation.

Procedures. Participants were divided into two groups. Two separate versions of videotaped subordinate performance were prepared. Both versions showed subordinates one scene from all 64 situations. If one group saw the effective scene for a situation, the other group saw the corresponding ineffective scene. The result was that each group saw 16 different examples of performance for each of the 4 subordinates. Over both groups all 128 scenes were shown.

Both groups of participants participated in the full simulation which included the in-basket exercises and exposure to the videotaped presentation of subordinate performance. The in-baskets materials and videotapes were shown over four separate sessions, one per week for 4 weeks. This allowed for evaluation of the entire simulation under the most stringent experimental conditions. After the fourth session, subjects used the BARS scale to rate the performance of all 4 subordinate on each of the 4 performance dimensions.

Three alternative versions of the true scores were prepared for comparison with MBA student ratings. The first true score was based on the ratings intended when each scene was prepared. In this case, for each scene showing effective performance, the contribution of the scene to overall subordinate true score was considered a 6. For each scene showing ineffective performance, the contribution to the true score was a 2. To develop a true score for the subordinate on a performance dimension, the average of the four scenes shown which reflected on that dimension was used. For example, if the group saw 3 effective scenes with a value of 6 and 1 ineffective scene with a value of 2, the true score for that dimension would be 5. This version of the true score is referred to as the "intended" true score.

The second true score was developed in the same way but substituting the actual mean ratings from expert judges for the true score component rather than the intended score. Referring back to Table 2-1, if the subordinate performance shown to the group included the effective scene from situation 1 (i.e. scene 1a) the true score contribution to the performance of that subordinate was 5.6 rather than 6 as in the intended true score version. Again, the true score for a subordinate on the performance dimension was the average of the expert judge's rating for the four scenes. This version of the true score is referred to as the "expert" true score.

Finally, a third true score was created that used all the ratings provided by expert judges including any ratings that were not on the target dimension or for the target subordinate. Occasionally one or more judges' ratings indicated that a particular scene reflected on more than one performance dimension or on a subordinate who happened to be in the scene but was not the target subordinate. To determine whether these extra ratings provided important information about true differences in subordinate performance, a true score version was developed that included any rating provided by at least two judges. This true score considered the possibility that additional information was being provided on rating performance beyond that intended when the script was developed. Again, all applicable scores were averaged to provide a true score for the subordinate on the dimension. This version of the true score is referred to as the "all ratings" true score.

Results. Mean ratings for all participants who observed a particular set of scenes were compared to the three alternative versions of true scores for that combination of scenes. Each of the two groups had a seen a separate set of scenes that contributed to 16 different overall true scores (ratings of the 4 subordinates on 4 performance dimensions). The overall mean ratings for both groups was then correlated with the three alternative versions of that true score. In other words, the means for both group's 16 ratings were correlated with the three alternative versions of the trues scores for the scenes they had observed. The results are presented in Table 2-2.

TABLE 2-2
CORRELATIONS BETWEEN STUDY B GROUP RATINGS
AND ALTERNATIVE TRUE SCORES VERSIONS

Variable	1	2	3	4
1. Group Ratings	1.00			
2. Intended Score	.68**	1.00		
3. Expert Score	.71**	.97**	1.00	
4. All Rating Score	.68**	.97**	.90**	1.00

Note: Two-tail significance
** p<.01

Each alternative true score definition was highly correlated. Additionally, each version of the true score was closely related to the mean of the ratings provided by subjects who had observed those versions of subordinate performance (intended true score *r* .68, expert rating true score *r* .71, all rating true score *r* .68).

Conclusions. The above analysis supports two important conclusions. First, the managerial simulation is an effective instrument in meaningfully manipulating subordinate performance in a realistic yet carefully controlled laboratory setting. Study B participants were exposed to 2 different versions of 4 subordinates' performance across 4 different performance dimensions and the average correlations between intended true score and subjects ratings was .70. This suggests that using different videotaped scenes to create different profiles of performance across 4 subordinates and 4 performance dimensions is an effective means of manipulating subordinate performance in meaningful ways.

The second important conclusion is that all three alternative versions of the true score carry very similar information about the subordinate performance depicted in the videotape and as a result, any of three versions is acceptable. The intended true score version provides the simplest definition of the true score and a review of the correlations between the alternative versions of the true score suggests there is little to be added by using a more complex measure of the true score. Use of the intended true score version is also supported by information from Study A which showed that independent expert judges consistently rated the target individual and dimension in the intended direction and the mean value of their ratings closely approximated the intended value of 6 for effective scenes and 2 for ineffective scenes.

Relationship to the current study

The methods described above were used in the current study to develop a true score of subordinate performance information presented to subjects. Each scene which showed a subordinate performing effectively on a performance dimension was assigned a true score value of 6 and each scene shown that presented ineffective performance information was assigned a value of 2. The overall subordinate true score on a performance dimension was calculated by taking the mean performance value of the 4 scenes used to present that subordinate's performance. This technique assumes that raters should give equal weight to each piece of performance information. This technique is justified when using the scenes developed above for the following reasons. First, each scene was carefully constructed to depict a subordinate behaving in a way that closely resembled one of the behavioral anchors used on each dimension of the BARS scale that subjects used to make their performance ratings. Since the scales provide several behaviors that reflect on a given level of performance, it is difficult to determine why one behavior should be assigned more weight than another. Second, judges in the Study A described above evaluated the scale value of each specific behavior using the BARS scale. Since their ratings closely approximated the intended values of each behavior, this can be taken as the true contribution of that behavioral episode to the overall level of subordinate performance. The remainder of this chapter describes the procedures and methods used in this study.

Overview of Procedures Used in This Study

This section provides an overview of the procedures followed in conducting this study. Special versions of the managerial simulation consisting of in-baskets and videotaped presentation of subordinate performance were developed from the material described above.

Procedures

The experimental portion of the study was conducted over a 2-week period. Participants were required to attend two 2-hour sessions, one during each of the two weeks. During both of these sessions students were given an in-basket exercise. The first week's in-basket familiarized subjects with their roles and responsibilities and presented a series of challenging administrative tasks. During the second week's session a different in-basket was provided which placed participants in the same situation but at a later time (two weeks later). Subjects were encouraged to provide detailed responses to all in-basket items.

The videotaped scenes used to present subordinate performance information were interspersed throughout both weeks of the simulation with approximately half of the scenes shown the first week and the other half shown the second week. Subjects were allowed to take notes on performance information presented through the videotaped scenes. They were unaware that they would not be allowed to use these notes when making their performance evaluations.

Following the second session students completed the performance ratings, observational accuracy check, manipulation check, and the simulation reaction check. These forms are described below in the instruments section.

Participants were required to complete a written assignment based on the treatment condition to which they were assigned. This assignment was due one week after the end of the second session. The study protocol is located in Appendix D.

Analysis

Test of hypotheses were conducted using t-tests to test for differences between group scores on key dependent variables. Additionally, a series of 2 X 2 analyses of variance (ANOVA) considered differences between treatment conditions. This focused F-test (1 degree of freedom in the numerator) compared differences in group means. This type of testing allows for precise consideration of the effects of manipulated variables on dependent variables and follows the reasoning outlined by Rosenthal and Rosnow (1985).

Subjects

Two hundred forty-seven undergraduates enrolled in an introductory course in management participated in this research and received course credit for their participation. The sample consisted of 136 men and 111 women with a mean age of 21.06 years. Most subjects were either in their third or fourth year of undergraduate work with a mean full time work experience of 19.78 months.

Male and female subjects were randomly distributed into 8 different treatment conditions to control for possible effects due to the sex of participants. As a result, there were 16-18 men and 13-15 women in each condition.

Manipulations of Accountability and Rating Outcomes

In this study both the participant's level of accountability and available information about possible outcomes that could result from their ratings were manipulated as described below.

Each participant was assigned to one of two accountability conditions:

1. ACCOUNTABLE for justifying their performance evaluations
2. NOT ACCOUNTABLE for justifying their performance evaluations

Subjects in the accountable condition were told that they were required to provide a written justification for their performance ratings. Subjects in the not accountable condition were told their ratings will remain anonymous and that they would be required to complete a written critique of the overall simulation. Subjects in both conditions were told that they will receive up to 5 points of extra credit for their participation and that the amount of extra credit would depend on their performance in the simulation and on their written assignment. Copies of the accountability manipulation are included in Appendix E.

Subjects were also assigned to one of four performance appraisal outcome conditions. In these conditions, subjects were provided information about the outcomes or effects of previous performance rating decisions on the organization. These outcome manipulations were presented to subjects in each treatment group through a letter from their "supervisor" presented to them just prior to presentation of the simulation. Additionally, subjects in the first three rating outcome situations were provided letters or memos within the contents of their in-baskets that reinforced the described manipulations. These materials are located in Appendix F. A summary of each outcome situation is provided below.

1. INFLATIONARY OUTCOME - The supervisor's letter said that the subordinates in the subject's division have consistently been rated lower than peers in other divisions and as a result, there have been no promotions from the subject's division in the last 5 years.

2. EQUITABLE TREATMENT OUTCOME - The supervisor's letter expressed concern over the low ratings given to women in the organization. The supervisor also discussed a pending lawsuit over the low ratings.

3. ACCURACY OUTCOME - The supervisor's letter expressed concern over how the consistent ratings inflation within the organization has made it impossible to use performance evaluations for administrative purposes (promotion, job selection etc.). As a result, subjects in this condition were encouraged to provide ratings that accurately reflected true differences of subordinates within each performance dimension.

4. COMPARISON GROUP - The supervisor's letter provided no information about the effects of previous performance evaluations. It described a generic list of the major challenges facing the organization.

Subordinate Performance Presented In This Study

The videotaped scenes of subordinate performance described earlier provided an large number of subordinate true score performance profiles that could be presented. Profiles for this study were constructed to facilitate tests of the hypotheses discussed in Chapter 1 and the rating outcomes described above. As a result, there were three criteria for the subordinate profiles. First, the true score for the profile of the highest performing subordinate on a dimension was developed to approximate the average score for subordinates in other divisions as presented in the manipulation given to subjects in the inflationary outcome condition (i.e a rating of 6). The true score for the other subordinates was lower than the average shown to people in the "inflationary condition". This difference was designed to create pressure on participants to inflate the ratings of their subordinates so they would be similar to ratings assigned to subordinates in other divisions.

Second, the overall true score for women was created to be lower than the true score for males on some performance dimensions. This difference was designed to create pressure on participants in the "equitable treatment" condition to artificially inflate the ratings of women to comply with pressure to provide equivalent ratings to male and female subordinates.

A third criterion was to provide variability in the overall strengths and weaknesses in each subordinate. This situation emphasizes the need for subjects to pay careful attention to within subordinate performance variation on individual performance dimensions. Table 2-3 presents the true score of simulation subordinates based on scenes selected for use in this study to meet the criteria described above.

TABLE 2-3
SUBORDINATE TRUE PERFORMANCE SCORES

	ADMIN	WORK EFFORT	COORD	SUPERVIS	OVERALL
ALICE	6	3	5	4	4
BILL	6	4	6	6	5.33
CAROLE	6	5	3	3	3.66
DAVID	6	6	4	5	4.5

Since equivalent performance information is presented for all subordinates on the administrative dimension, the overall true score was based on the average of the true score values of the work effort, coordination and supervision dimensions.

Instruments

Rating Format

Individuals used the BARS scale specially designed for the jobs depicted in the in-basket/videotapes. This scale was discussed earlier and can be found in Appendix B. Subject ratings were collected using the rating form found in Appendix G.

Demographic Questionnaire

Subjects answered demographic questions prior to the beginning of the experiment. The questions pertained to background information including age, sex, and work experience. A copy of this questionnaire is included in the Appendix H.

Manipulation Check

After the final session, subjects answered questions dealing with the context of the rating including specific questions about the information they received in the simulation. Subjects were provided with a list describing each of the materials prepared to manipulate the different outcome conditions and were asked to identify which materials they recalled receiving. These results were used to test that each group recalled receiving material unique to their condition. Coding for this variable is discussed in Chapter 3. Copies of the manipulation check are included in the Appendix I.

Observational Accuracy Check

This form assessed participant recall of specific behavioral events presented including whether the event presented positive or negative information about subordinate performance. It listed 32 statements that were either accurate and inaccurate descriptions of the scenes shown during the simulation. Subjects were asked to identify which scenes actually occurred. Copies of the observational accuracy instrument are included in Appendix J.

Post-experiment Questionnaire

This form assessed subject reactions to the entire simulation. It included a variety of questions to assess whether subjects reactions differed by treatment condition. Question responses were used as self-report information about subjects attitudes and reactions towards the tasks presented in the simulation. A copy of this questionnaire is located in Appendix K.

Dependent Variables

Variables for Testing Hypotheses

Performance ratings

Performance ratings were collected after the second session. Subjects rated the four subordinates, Alice, Bill, Carole, and David on the four performance dimensions. Since only behavior on three of the dimensions varied between subordinates, those ratings were used for computation of the differential accuracy measure described below. The mean ratings were also used to assess the overall favorability of ratings assigned by subjects in different treatment conditions. Subjects were told to record a rating from 1 to 7 for each subject with 7 being the most favorable rating.

Differential Accuracy

Differential accuracy measures the ratee by dimension interaction and was used to test predictions of differences in accuracy between treatment conditions. Differential accuracy scores indicate how well the rater can differentiate between different ratee performances over different performance dimensions. This measure were computed using the equation presented by Cronbach (1955).

Observational Accuracy

This measure indicates how accurately participants recalled specific performance information. It is the sum of the number of performance episodes the participant correctly identifies as having observed out of 32 episodes described in this instrument. If a subject correctly identifies whether or not he or she saw all 32 of the episodes described they would receive a score of 32. Internal reliability of this instrument was calculated using Cronbach's alpha. This resulted in an reliability estimate of .82.

Variables for Additional Analyses

Attentiveness

This measure indicates how attentive participants were to the presentation of performance information. Two judges, blind to the subject's treatment condition, rated subjects attentiveness using the attentiveness rating scale located in Appendix L. One judge made his ratings during the first session and the second judge made his ratings during the second session. Judges observed the subjects while they worked on their in-baskets and watched the videotaped scenes of subordinate performance. Judges were strategically positioned within the room so they could view participant actions during the presentation of the performance information. Judge's assigned each participant a rating from 1 to 3. Inter-rater reliability was determined by calculating the intra-class correlation between the two sets of ratings. This correlation was adjusted using the Spearman-Brown correction formula to determine that the reliability of the composite of the two judges' ratings was .67. The ratings of the two judges were summed to form the participant's overall attentiveness score.

Note Taking

This measure provides a judgement of the quality and quantity of subject's notes taken voluntarily in response to performance relevant behavior presented in the simulation. While subjects were provided paper for taking performance related notes, they were not encouraged or discouraged from taking them. Subjects were unaware that they would not be able to use these notes when making their performance evaluations. The performance related notes from both sessions were combined and provided to independent judges blind to the subject's treatment condition. These judges independently rated both the quality and quantity of notes taken using the note taking rating scale located in Appendix M. Judges assigned a rating of from 0 to 4 to the combined set of notes. Inter-rater reliability was determined by calculating the intra-class correlation between the two sets of ratings. This correlation was adjusted using the Spearman-Brown correction formula to determine that the reliability of the composite of the two judges' ratings was .92. The ratings of the two judges were summed to form the participant's note taking score.

Self-report measures

Two scales were developed from subject reactions to the post-experiment questionnaire located in Appendix J. These scales were used to support additional analyses found in Chapter 3. Cronbach's alpha is reported as a measure of internal consistency of each variable.

Engagement. One variable measured the level of engagement felt by subjects toward their participation in the simulation. It was felt that subjects who were more cognitively engaged in the simulation would report spending more time thinking about the simulation and specific challenges presented in the simulation. This variable is based on subject responses to items l, m, and n (alpha .69) of the "Post-experiment Questionnaire".

Concern. Finally, a single item scale, based on item o of the "Post-experiment Questionnaire". This measure was used to assess how concerned the subject was that they may not receive full credit for participation in the simulation.

CHAPTER 3
RESULTS

This chapter presents results of the manipulation checks, tests of hypotheses, and additional analyses.

Manipulation Checks

Accountability Manipulation

The manipulation check for the accountability manipulation was simply whether subjects demonstrated that they understood the requirements of the final assignment by turning in the proper assignment. Subjects in the accountable condition justified their performance ratings. Subjects not held accountable for their ratings prepared a written critique of the simulation. All subjects submitted the correct final assignment based on the accountability manipulation.

Rating Outcome Situational Manipulation

After the second week's session subjects were shown a list which described every letter or memo used to manipulate the rating outcome conditions. Each letter described was seen by subjects in only one of the rating outcome conditions. Subjects were asked to identify which of the materials they recalled receiving by placing a 1 if they recalled the letter and a 0 if they did not recall the letter. Seven letters were described. Subjects in the inflationary, equitable treatment, and accuracy outcome conditions had been exposed to two of the letters, subjects in the comparison group had seen only one of the letters.

Four variables were constructed which consisted of the sum of each subject's score on the letters that were related to each situation. For example, if a subject reported that he or she had seen both of the letters that were presented to one of the outcome conditions, they would receive a 2 for that set of materials. If they reported they hadn't seen either of those letters, they would receive a 0 for that condition. Ideally, a subjects assigned to the inflationary outcome condition would receive a 2 for materials presented to that condition and a 0 for materials presented to the other 3 conditions. Table 3-1 shows the group means for each relevant outcome condition for the variable comprised of responses to the description of material provided to that condition. Means for all subjects assigned to other groups who did not receive those materials is also provided.

The average score for subjects assigned to the inflationary outcome condition on materials they had been shown was $\underline{M} = 1.38$. If each member of this group had correctly identified that they had received the letters that made up that outcome manipulation, the group's mean score

TABLE 3-1
MANIPULATION CHECK RESULTS

	Number of Items Reported By					
Material	Relevant Experimental Groups			All Other Groups		
	N	Mean	SD	N	Mean	SD
Inflationary	63	1.38	.55	184	0.47	.60
Equitable Treatment	61	1.61	.59	185	0.21	.44
Accuracy	63	1.44	.50	184	0.46	.60
Comparison Group	60	0.92	.18	186	0.50	.50

would have been a 2. Mean score on the inflationary outcome materials for subjects assigned to the other 3 conditions was .47. If all members of those conditions had correctly recalled that they had not seen the inflationary outcome letters the group mean score would be 0. Because only one letter was shown to subjects in the comparison group, the score for subjects in that condition should be a 1 and the score for subjects assigned to the other conditions should be a 0.

One-tailed t-tests were used to determine if the subjects in groups that had been shown letters designed to present a specific manipulation answered yes to seeing those letters more consistently than subjects in groups who had not been shown the letters. All t-tests showed significant differences between groups ($p<.001$) in the materials they recalled receiving. Therefore, despite receiving over 100 pages of memos and letters within the two in-baskets, this analysis shows that subjects who were provided materials designed to present a certain manipulation recalled receiving those materials more consistently than subjects in the other conditions.

Methods of Analysis

Independent sample t-tests are used to consider hypothesized differences between groups. Additional analyses included a series of 2 x 2 analysis of variance (ANOVA) to support a focused look at the effects of each rating outcome condition when compared to the comparison group. Finally, LISREL is used for structural equation modeling and analysis.

Tests of Hypotheses

Independent sample t-tests were used to test hypotheses presented in Chapter 1. Each test made one of two comparisons depending on the hypothesis being tested. One group of t-tests compared subjects in selected cells with non-accountable subjects in the comparison group. These latter subjects were not held accountable for their ratings and were not provided a desirable rating outcome manipulation. As a result, they can be considered a control group for comparison purposes. A second group of t-tests contrasted differences between accountable subjects and non-accountable subjects within a specific rating outcome condition.

Consistent with the directional predictions stated in each hypothesis, all t-test are one-tailed and considered significant when the probability of a type 1 error for the resulting t-value was less than or equal to .05. The t-tests used to test specific hypotheses are considered "protected" against the possibility of capitalizing on chance if they were either planned before the data was collected or they are supported by the overall F computed as part of the ANOVAs discussed later in this section. This method is consistent with arguments made by Rosenthal and Rosnow (1984). For additional discussion see Ballam (1963).

Rating accuracy

The first set of hypotheses considered differences in rating accuracy that result from different treatment conditions. Table 3-2 presents cell means for each condition based on the dependent variable of differential accuracy. Because differential accuracy is a measure of rating error, subjects with a higher score on the measure had more error and were therefore less accurate.

TABLE 3-2
MEAN VALUES AND STANDARD DEVIATIONS FOR
DIFFERENTIAL ACCURACY BY TREATMENT CONDITION

	Inflation		Equitable Treatment		Accuracy		Compare Group		Overall	
Accountability	MEAN	SD	MEAN	SD	MEAN	SD	MEAN	SD	Mean	SD
Yes	1.03	.55	1.20	.51	.91	.53	1.00	.74	1.04	.59
No	1.38	.64	1.52	.80	1.23	.63	1.34	.65	1.37	.68
Overall	1.20	.61	1.36	.68	1.07	.60	1.17	.71		

Hypothesis 1

Hypothesis 1 predicts that accountable raters would be more accurate than non-accountable raters except when desirable rating outcomes could be achieved by providing inaccurate ratings. The differential accuracy obtained by accountable subjects in the comparison group was compared to that of non-accountable subjects. As shown in Table 3-2, accountable subjects were more accurate (*M* = 1.00, *SD* = .74) than non-accountable subjects (*M* = 1.34, *SD* = .65) (*t*(58) = 1.84, *p*<.05).

Hypothesis 1 was also tested by comparing the differential accuracy obtained by accountable subjects in the inflationary and accuracy outcome conditions with their non-accountable counterparts. In the inflationary group, accountable subjects were significantly more accurate (*M* = 1.03, *SD* = .55), than non-accountable subjects (*M* = 1.38, *SD* = .64), *t*(61) = 2.29, *p*<.05). The predicted differences in accuracy was also found in the accuracy condition. Accountable subjects were significantly more accurate (M = .91, SD = .53), than non-accountable subjects (M = 1.23, SD = 1.23) ($t(61) = 2.24$, $p<.05$). Hypothesis 1 was supported in all three conditions where accountable subjects were predicted to be more accurate than non-accountable subjects.

Hypothesis 2

Hypothesis 2 predicts that accountable subjects in the equitable treatment outcome condition, who were encouraged to rate male and female subordinates equally, would be less accurate than non-accountable subjects in the comparison group. This analysis does not support the hypothesis. In fact, the pattern of means shown in Table 3-3 means suggests that accountable subjects in the equitable treatment condition were more accurate (M = 1.20, SD = .51) than non-accountable subjects in the comparison group (M = 1.34, SD = 1.52). Because one-tailed tests are used in this analysis, this difference in means is not interpretable.

Hypothesis 3

Hypothesis 3 predicts that accountable subjects presented with a rating context where accurate ratings are desired would be more accurate than subjects in the comparison group regardless of accountability status. This hypothesis was partially supported. Accountable subjects in the accuracy treatment group (M = .91, SD = .53) were significantly more accurate than non-accountable subjects in the comparison group (M = 1.34, SD = .65) ($t(60)$ = 2.88, $p<.01$). However, while the patterns of means shown in table 3-2 suggests that their mean differential accuracy was lower (i.e. less error) than accountable subjects in the comparison group (M = 1.00, SD = .74), that difference was not significant ($t(60)$ = .62, p ns).

Observational accuracy

Table 3-3 presents scores for observational accuracy separately for each condition.

TABLE 3-3
MEAN VALUES AND STANDARD DEVIATIONS FOR
OBSERVATIONAL ACCURACY BY TREATMENT CONDITION

	Inflation		Equitable Treatment		Accuracy		Compare Group		Overall	
Accountability	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Yes	25.7	3.3	24.4	2.7	24.1	3.0	24.1	3.1	24.9	3.1
No	23.1	4.4	24.1	2.9	23.1	3.2	23.8	3.7	23.5	3.6
Overall	24.5	4.1	24.3	2.8	24.1	3.2	24.0	3.4		

Hypothesis 4

Hypothesis 4 predicts that accountable subjects would have more accurate recall of performance episodes than non-accountable subjects. Within the comparison group, the observational accuracy of accountable raters (M = 24.1, SD = 3.1) was not significantly greater than that of non-accountable subjects (M = 23.8, SD = 3.7) ($t(58)$ = .34, ns).

However, in the inflationary outcome and the accuracy outcome conditions accountable subjects were significantly more accurate than subjects not held accountable for their performance ratings. Within the inflationary condition accountable subjects more accurately (M = 25.7, SD = 3.3) recalled performance episodes than non-accountable subjects (M = 23.1, SD = 4.1) ($t(61) = 2.74$, $p<.01$). Within the accuracy treatment condition accountable subjects also more accurately (M = 24.1, SD = 3.1) recalled more performance episodes than non-accountable subjects (M = 23.1, SD = 3.2) ($t(61) = 2.54$, $p<.01$).

Hypothesis 4 was supported in two of the three conditions where subjects were not placed in a rating context where desirable outcomes could be achieved by making inaccurate ratings.

Hypothesis 5

Hypothesis 5, predicts that accountable subjects placed in a rating context where a desirable outcome could be achieved by rating female subordinates higher than true level of performance warrants, would have lower observational accuracy than non-accountable subjects. It was predicted that these raters would be susceptible to contextual pressures and adopt an information search strategy in which they attended to only positive information. However, because the analysis of hypothesis 2 found that raters in this condition were not less accurate than raters in the comparison group, that suggests that these raters should also have a better recall of performance episodes. Consistent with results found for hypothesis 2, hypothesis 5 was not supported either. In fact, the pattern of means shown in Table 3-3 means suggests that accountable subjects in the equitable treatment condition scored higher on the measure of observational accuracy (*M* = 24.4, *SD* = 2.7) than non-accountable subjects in the comparison group (*M* = 23.8, *SD* = 3.7). Again, because one-tailed tests are used in this analysis, this difference in means is not interpretable.

Favorability of Mean Ratings

Table 3-4 presents overall means for each separate condition based on the overall mean ratings assigned by subjects.

TABLE 3-4
MEAN VALUES AND STANDARD DEVIATIONS FOR
MEAN RATINGS BY TREATMENT CONDITION

	Inflation		Equitable Treatment		Accuracy		Compare Group		Overall	
Accountability	MEAN	SD	MEAN	SD	MEAN	SD	MEAN	SD	MEAN	SD
Yes	5.0	.40	4.8	.43	4.8	.45	4.7	.46	4.8	.44
No	4.7	.39	5.0	.50	4.7	.58	4.7	.41	4.8	.49
Overall	4.9	.42	4.9	.48	4.7	.52	4.7	.43		

Hypothesis 6

Hypothesis 6 considered the effect of the inflationary outcome manipulation and predicted that accountable subjects in this situation would provide more favorable ratings of subordinate performance in response to contextual pressure to make their subordinate's ratings more competitive with subordinates in other divisions. As predicted, accountable subjects in this condition made more favorable ratings (M = 5.0, SD = .40) when compared to non-accountable subjects in the comparison group (M = 4.7 SD = .41) ($t(61) = 2.73$, $p<.01$). These accountable subjects also rated subordinates more favorably than subjects in the same condition who were not held accountable for their ratings (M = 4.7 SD .39) ($t(61) = 2.73$, $p<.01$).

As mentioned during the discussion of this hypothesis, previous research has suggested that raters can inflate ratings without significantly reducing differential accuracy. This analysis supports that conclusion. Accountable subjects did rate subjects more favorably, but they did so in a way that still recognized true differences in performance between subordinates and across diverse performance dimensions. If, as these results suggest, they equally inflated all ratings to the same degree, they were not substantially penalized on a measure of differential accuracy.

Hypothesis 7

Table 3-5 presents overall means for each separate condition based on the overall mean ratings assigned by subjects to their female subordinates.

TABLE 3-5
MEAN VALUES AND STANDARD DEVIATIONS FOR
FEMALE MEAN RATINGS BY TREATMENT CONDITION

	Inflation		Equitable Treatment		Accuracy		Compare Group		Overall	
Accountability	MEAN	SD	MEAN	SD	MEAN	SD	MEAN	SD	MEAN	SD
Yes	4.1	.71	3.7	.64	3.9	.68	3.7	.70	3.84	.69
No	3.7	.57	4.3	.74	3.6	.73	3.8	.51	3.83	.69
Overall	3.9	.42	4.0	.48	3.7	.52	3.7	.43		

Subjects in the equitable treatment outcome condition were exposed to a rating context where a desirable outcome could be achieved by rating male and female subordinates equally. Hypothesis 7 predicts that only accountable raters in this condition would be affected by this manipulation and would rate their female subordinates more favorably. The results of earlier hypotheses have shown that accountable subjects in this condition were not less accurate and did not have poorer recall of specific behavioral events. Consistent with this, hypothesis 7 was also not supported. Contrary to prediction, mean ratings of female subordinates were not more favorable for accountable subjects in the equitable treatment outcome condition (M = 3.7, SD = .64) than for non-accountable subjects in the comparison group (M = 3.8, SD = .51). The use of one-tailed tests again makes this difference in means not interpretable.

Dilution Effect

A final research question presented in Chapter 1 dealt with the dilution effect, the possibility that accountability may lead subjects to give weight to irrelevant performance information. Since subjects who were accountable were more accurate on measures of observational accuracy and differential accuracy, there is a strong implication that the performance information gathered by accountable subject was not diluted by irrelevant information about subordinate work performance.

Additional data gathered for this study supports this implication. Two dependent variables were constructed that summed the total number of positive and the total number of negative events recalled by subjects. The assumption behind creation of these variables was that if accountable subjects had attended to and weighted irrelevant information, they would subsequently recall a higher number of events when they made their performance ratings.

Results of t-tests show there was not a significant difference between the number of positive events recalled by accountable subjects (M = 17.8, SD 6.96) as compared to subjects not held accountable (M = 16.5, SD = 6.50), t(243)

= 1.51, _p_ = .133). There were also no significant differences on the number of negative scenes reported by accountable subjects (_M_ = 11.5, _SD_ 3.62) as compared to non-accountable subjects (_M_ = 10.8, _SD_ = 4.29), _t_(243) = 1.37, _p_ = .171). These results do not support a dilution effect.

Summary of t-test results

The results of t-tests of hypotheses suggest that the accountability manipulation led to improved accuracy. Across three distinct rating contexts, accountable subjects rated their subordinates more accurately than their non-accountable counterparts. When accountable subjects were required to provide ratings they could not justify, as in the case of artificially inflating female ratings relative to males, they resisted the situational pressure to do so.

Within the inflationary and accuracy outcome conditions, accountable subjects recalled information about specific incidences of subordinate performance more accurately than non-accountable subjects. However, in the comparison group, accountable subjects did not have significantly more accurate recall than non-accountable raters. These results suggest that holding raters accountable may lead raters to attend to performance information in a way that makes them have better recall of specific performance events.

Additional Analyses

2 X 2 ANOVA Results

Whereas the t-tests of hypotheses compared results between selected conditions, they did not directly test for the overall effects of different rating outcome conditions. As a result, additional analyses using ANOVA allowed for a focused test of the contrast differences between subjects in each rating outcome condition with subjects in the comparison group.

Inflationary outcome condition

Table 3-4 presents the results of a 2 X 2 ANOVA considering the effect of a rating context where a desirable outcome can be achieved by inflating the rating of all subordinates as compared to the comparison group.

The ANOVA results reported in Table 3-6 shows that there was an interaction between accountability and rating outcome ($F(1,122) = 3.69$, $p < .05$). Based on means reported in Table 3-4, it appears that only when raters were held accountable were they influenced by contextual pressure to inflate their ratings. A Duncan's multiple range test reported significant ($p<.05$) differences between the mean ratings of accountable raters in the inflationary outcome condition and the other three conditions contrasted in this analysis. This supports the prediction that raters will be

TABLE 3-6
RESULTS OF 2X2 ANOVA COMPARING THE INFLATIONARY OUTCOME CONDITION WITH COMPARISON GROUP

Source of Variation	Mean Rating	
	F	Signif
Accountability	4.47	.037*
Rating Context	2.92	.090
Interaction AxR	3.69	.050*

Note: df = 1, 122
*p ≤.05 ** p ≤.01

influenced to achieve specific rating outcomes when they are linked to their ratings by accountability.

Equitable treatment condition

Table 3-7 presents the results of a 2 X 2 ANOVA considering the effect of a rating context where a desirable outcome can be achieved by inflating the ratings of female subordinates to insure they are equal to ratings of male subordinates.

Table 3-7 shows a significant interaction between accountability and rating context. It was predicted that accountable raters in the equitable treatment condition would rate female subordinates more favorably than raters in the comparison group. However, as suggested, the pattern of mean ratings for females found in Table 3-5 indicates that non-accountable raters gave more favorable ratings to female subordinates. Based on the significant interaction for

TABLE 3-7
RESULTS OF 2X2 ANOVA COMPARING THE EQUITABLE TREATMENT OUTCOME CONDITION WITH COMPARISON GROUP

Source of Variation	Female Mean Rating	
	F	Signif
Accountability	6.61	.011*
Rating Context	5.21	.024
Interaction AxR	4.76	.031*

Note: df = 1, 120
*p ≤.05 ** p ≤.01

female means ($F(1,120) = 4.76$, $p<.05$), a Duncan's Multiple Range test was conducted to test for significant differences between conditions. These results revealed significant differences ($p<.05$) between the favorability of rating provided by non-accountable subjects in the equitable treatment condition and subjects in the other three conditions considered in this ANOVA.

This result is interesting because it is contrary to other findings in this study. In the equitable treatment condition, subjects who were not held accountable for their ratings rated female subordinates more favorably when presented with a rating context where a desirable outcome was to insure the performance ratings of female subordinates equaled those for males.

Accuracy treatment condition

Table 3-9 presents 2 X 2 ANOVA results comparing the differential accuracy of subjects in the comparison group with subjects in the accuracy outcome condition who were encouraged to provide more accurate ratings to facilitate organizational decision making.

TABLE 3-8
RESULTS OF 2X2 ANOVA COMPARING THE ACCURACY
OUTCOME CONDITION WITH COMPARISON GROUP

Source of Variation	Differential Accuracy	
	F	Signif
Accountability	8.13	.005**
Rating Context	.80	.374
Interaction AxR	.00	.991

Note: df = 1, 122
*p ≤.05 ** p ≤.01

The overall significant effect of accountability on differential accuracy has already been discussed. This analysis shows there was no significant effect of rating context and no interactions.

Summary of 2 X 2 ANOVA results

The results of contrast analyses show that only accountable raters who were placed in the rating context where they could achieve a desirable outcome by inflating their ratings showed significant influences on the quality of their ratings. Accountable raters in the equitable treatment group were not affected by a rating context where a desirable rating outcome could be achieved by rating female subordinates more favorably. Similarly, while accountable subjects in the accuracy outcome condition had the lowest mean rating on differential accuracy across all condition, their ratings were not significantly more accurate than accountable raters in other conditions.

Structural Model Analysis

Results reported so far show that subjects who were required to justify their ratings recalled specific performance events more accurately (observational accuracy) and rated performance more accurately (differential accuracy). To explore the factors that contribute to this observed relationship, several process variables were considered. One self-report variable was constructed from student responses to the post-experiment questionnaire as discussed in Chapter 2. This variable measured the level of engagement felt by subjects toward their participation in the simulation. Subjects who were more cognitively engaged in the simulation were expected to spend more time thinking about the simulation and the specific challenges presented in the simulation.

In addition to the "engagement" variable, two behavioral variables were measured as discussed in Chapter

2. Attentiveness measured how attentive individual subjects were when subordinate performance information was presented via the television monitor. The quality and quantity of note taking was measured based on how useful voluntary notes taken by the subjects would have been if they had been used during the performance evaluations.

Table 3-9 presents means, standard deviations, and correlations for the accountability, observational accuracy, and differential accuracy as well as the self-report and behavioral variables.

TABLE 3-9
MEANS, STANDARD DEVIATIONS, AND CORRELATIONS
BETWEEN BEHAVIORAL AND SELF-REPORT VARIABLES

Variable	Means	SD	1	2	3	4	5	6
1. Accountability	1.49	.50	(1.00)					
2. Observational Accuracy	24.21	3.38	-.18	(.82)				
3. Differential Accuracy	.63	.59	.26**	-.27	(.65)			
4. Engagement	10.77	2.25	-.22**	.20**	-.40**	(.69)		
5. Attentiveness	3.87	1.14	-.19**	.47**	-.26**	.12	(.67)	
6. Note taking	4.80	1.99	-.34**	.35**	-.15*	.13*	.25**	(.92)

Two-tail significance. N = 247
* $p \le .05$ **$p \le .01$
Note 1: Accountability was coded a 1 for accountable subjects and a 2 for non-accountable subjects
Note 2: Differential accuracy measures the average level of error in subject ratings. As a result, lower scores on the measure indicates increased accuracy.
Note 3: Reliabilities are reported in the diagonal.

Based on the correlation between variables shown in Table 3-9, structural equation modeling was conducted to explore the possibility that subjects required to account for their ratings responded to that pressure by adopting certain attitudes and behaviors relative to the rating task. These attitudes and behaviors might have contributed to improved observational accuracy and differential accuracy.

Rating Process Model

A model was developed which proposed that accountable subjects were more attentive to subordinate performance, took more notes about subordinate behavior, and were more engaged in dealing with the problems presented in the simulation. As a result, higher scores on these three variables would be related to improved observational accuracy which in turn would lead to more accurate performance evaluations. Based on these proposed relationships, variables were analyzed in the following causal order: first accountability; second attentiveness, note taking, and engagement; third observational accuracy; fourth differential accuracy.

Structural equation modeling was used to test the relationships between accountability, rating process variables, and both measures of accuracy. Structural equation modeling provides advantages in this situation because it allows for both assessment and modification of theoretical models (Anderson & Gerbing, 1988). Two significance tests were used to assess the structural relationships between variables included in the models.

First, a model which specified a direct path between each variable and all subsequent dependent variables was compared to a series of nested models where certain paths were restricted to zero using the sequential chi-square difference approach recommended by Anderson and Gerbing (1988). Second, the significance level of individual paths was used for exploratory purposes to help determine the best fitting nested model.

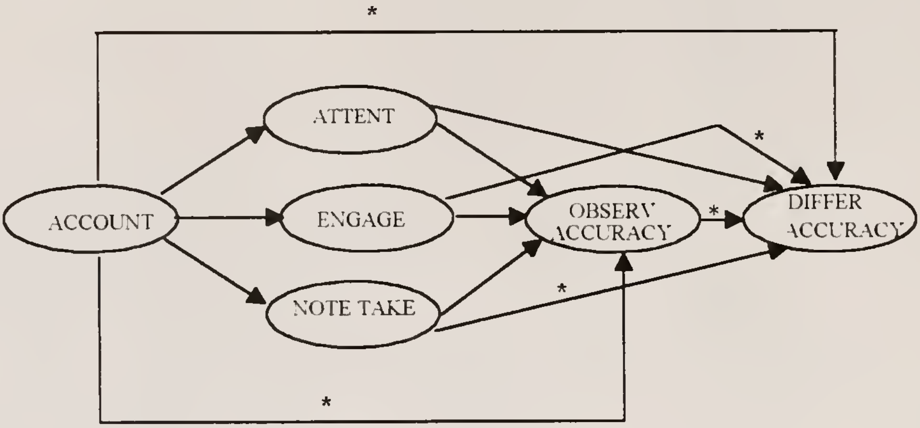
One advantage of structural equation modeling using LISREL 8.03 (Joreskog & Sorbom, 1993) is that it allows for significance testing of structural models after correcting for attenuation due to unreliability of the measures. In the following analyses reliability for the accountability variable was assumed to be 1. Reliability for differential accuracy is difficult to assess because each subject has only one score. However, that score actually represents the average error for the subject across their 12 ratings. Cronbach's alpha was estimated by treating the error in each rating as a single item and then computing the internal consistency of all 12 items ($\underline{a} = .65$). Reliability estimates for the remaining variables were presented in Chapter 2 and are shown in Table 3-9. Error terms were directly entered into the structural equations using the following formula: [1 - Reliability] X Variance.

The base model (Model 1) analyzed had direct paths from each variable to all subsequent variables in the model

(Figure 3-1). The chi-square test result was, X^2 (3,N=247) = 11.14, $p<.05$. A significant chi-square suggests that the model does not fit the data. However, additional measures of fit including the goodness-of-fit index, adjusted goodness-of-fit index, parsimonious goodness-of-fit index, and comparative fit index presented in Table 3-10 all suggest a good fit of the model to the data.

A second model (Model 2) was developed to test whether the three process variables fully mediated the relationship between accountability and accuracy. This model constrained the direct paths between accountability and both observational and differential accuracy to zero (Figure 3-2). The chi-square test result was, X^2 (5, N=247) = 12.42, $p<.05$. The fit indices presented in Table 3-10 show that the mediating model provided an improved fit over the base model. The chi-square difference test, as shown in Table 3-11, indicated that there was not a significant difference between the models.

A third model (Model 3) tested the proposed theoretical relationship between the variables (Figure 3-3). That is that accountability is related to the subject's attentiveness, engagement, and note taking which is positively related to observational accuracy which is in turn related to differential accuracy. The chi-square test result was X^2 (8, N=247) = 53.28, $p<.01$. There was a significant difference between Model 3 and Model 2 as shown



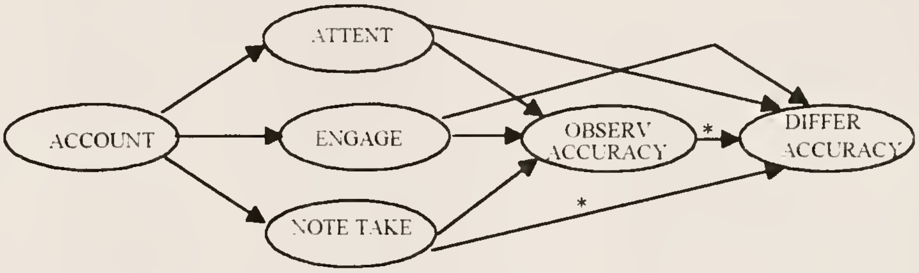
* indicates t-values for path coefficients are not significant at p<.05

FIGURE 3-1
MODEL 1 - BASE MODEL

TABLE 3-10
GOODNESS OF FIT INDICES FOR MODELS 1-7

Model	X^2	df	GFI	AGFI	PGFI	CFI
Model 1 Base Model	11.14	3	0.98	0.89	0.14	0.95
Model 2 Mediating Model	12.42	5	0.98	0.93	0.23	0.96
Model 3 Theoretical Model	53.28	8	0.94	0.83	0.36	0.75
Model 4 Revised Model	12.67	7	0.98	0.95	0.33	0.97
Model 5 Obs Accuracy Model	11.23	4	0.98	0.93	0.37	0.94
Model 6 Diff Acc Model A	12.25	4	0.98	0.93	0.26	0.93
Model 7 Diff Acc Model B	3.04	2	0.99	0.97	0.20	0.99

Note: GFI = goodness-of-fit index; AGFI= adjusted goodness-of-fit index; PGFI = parsimonious goodness-of-fit index; CFI = comparative fit index.



* indicates t-values for path coefficients are not significant p<.05

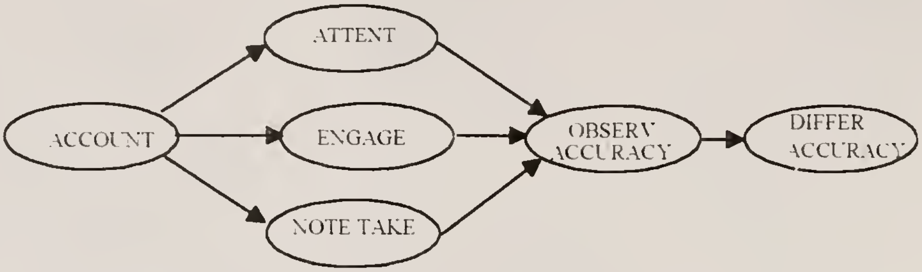
FIGURE 3-2

MODEL 2 - MEDIATING MODEL

TABLE 3-11
PARAMETER SIGNIFICANCE TESTS FOR MODEL COMPARISONS

			Model Comparison		
Model	X^2	df	Comparison	X^2	df
Model 1 Base Model	11.14*	3	--	--	--
Model 2 Mediating Model	12.42*	5	M_2-M_1	1.28	2
Model 3 Theoretical Model	53.28*	8	M_3-M_2	40.86*	3
Model 4 Revised Model	12.67	7	M_4-M_2	.25	2

* indicates chi-square statistics significant at p<.05

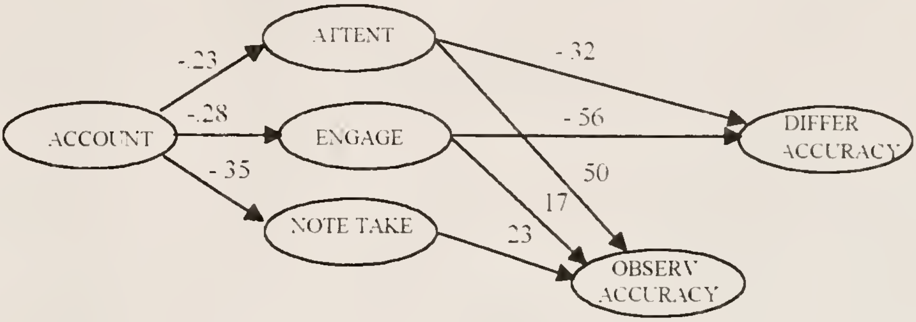


* all t-values for path coefficients are significant $p<.05$

FIGURE 3-3
MODEL 3 - THEORETICAL MODEL

in Table 3-11 and all fit indices showed that the theoretical model (Model 3) was a poorer fit to the data than the mediating model (Model 2).

At this point the t-values associated with the paths determined by the analysis of the mediating model were considered to determine if additional constraints were warranted. Two non-significant paths ($p<.05$) were deleted from the mediating model. The direct paths from note taking and observational accuracy to differential accuracy were constrained to be zero. Model 4, the revised model, is shown in Figure 3-4. The chi-square test resulting from this model was X^2 (7, N=247) = 12.67, p ns. The related fit indices shown in Table 3-10 also show that Model 4 provides the best fit to the data. Table 3-11 shows that the chi-



Note 1. All t-values for path coefficients are significant p<.05

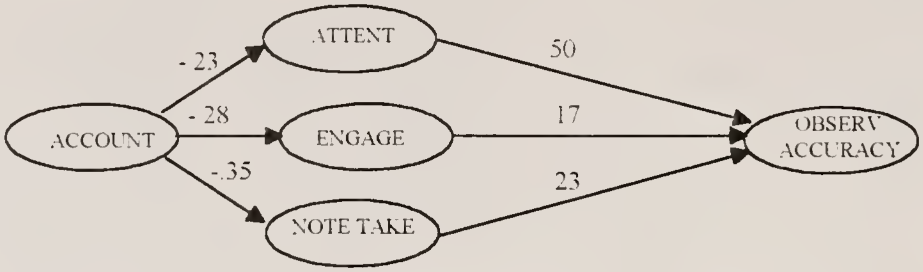
Note 2: Accountability variable was coded a 1 for accountable subjects and a 2 for non-accountable subjects. Lower values for Differential Accuracy indicates more accurate ratings.

FIGURE 3-4
MODEL 4 - REVISED MODEL

square differences between Model 2 and Model 4 was not significant.

Figure 3-4 also includes the standardized path coefficients for paths included in the revised model. These paths indicate that accountability had the strongest effect on note taking. Among the process variables attentiveness had a stronger effect on observational accuracy while engagement had the stronger effect on differential accuracy.

Figure 3-4 shows that observational accuracy was not related to differential accuracy when the three process variables were also included in the model. To simplify these relationships three additional structural models are presented. Figure 3-5 shows the relationship between those variables and observational accuracy (Model 5). The chi-



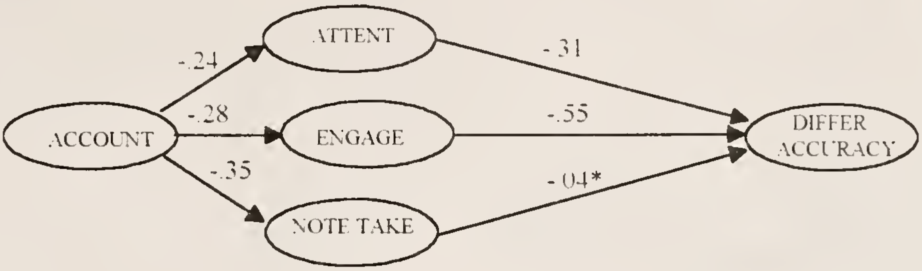
Note 1. All t-values for path coefficients are significant p<.05

Note 2. Accountability variable was coded a 1 for accountable subjects and a 2 for non-accountable subjects. Lower values for Differential Accuracy indicates more accurate ratings.

FIGURE 3-5
MODEL 5 - OBSERVATIONAL ACCURACY

square test resulting from this model was X^2 (4, N=247) = 11.23, $p<.05$. As shown in Table 3-10, while the chi-square test is significant, the remaining fit indices indicate the model is an acceptable fit to the data.

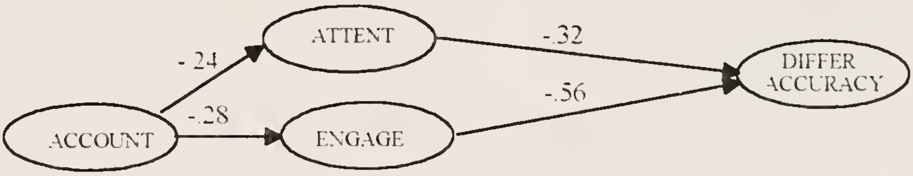
Figure 3-6 shows the relationship between accountability, the process variables, and differential accuracy (Model 6). The chi-square test resulting from this model was X^2 (4, N=247) = 12.25, $p<.05$. Figure 3-7 shows a similar model for differential accuracy with the effect of note taking removed (Model 7). The chi-square resulting from this model, X^2 (2, N=247) = 3.04, p ns, and the goodness of fit indices as shown in Table 3-10 show that this models provides a strong fit to the data.



Note 1. * indicates t-values for path coefficients are not significant p<.05.

Note 2. Accountability variable was coded a 1 for accountable subjects and a 2 for non-accountable subjects. Lower values for Differential Accuracy indicates more accurate ratings.

FIGURE 3-6
MODEL 6 - DIFFERENTIAL ACCURACY MODEL A



Note 1: All t-values for path coefficients are significant p<.05

Note 2 Accountability variable was coded a 1 for accountable subjects and a 2 for non-accountable subjects. Lower values for Differential Accuracy indicates more accurate ratings.

FIGURE 3-7
MODEL 7 - DIFFERENTIAL ACCURACY MODEL B

Summary of Structural Model Analyses

The squared multiple correlation coefficient was .37 for observational accuracy and .44 for differential accuracy. Additionally, these models showed that the effect of accountability on both observational and differential accuracy can be explained by variation in the process variables. It also shows that while note taking improved observational accuracy, it did not improve differential accuracy. Finally, observational accuracy did not account for variation in differential accuracy beyond that explained by the direct effect of attentiveness and engagement.

CHAPTER 4
DISCUSSION

Findings

The findings of this study show that differences in the rating context can lead to differences in rating quality. There are several conclusions which can be drawn from this research effort and perhaps just as importantly, several directions are suggested for future investigation.

Effects of Accountability on Rating Accuracy

Raters who were required to justify their performance ratings made more accurate ratings than their non-accountable counterparts regardless of the situational pressures presented by different possible outcomes of their ratings. Even in a rating context where desirable outcomes could be achieved by rating inaccurately, accountable raters still rated more accurately. It appears that the pressure of accountability prevented them from making a rating that was difficult to justify.

The results of the accountability manipulation are especially interesting in light of the fact that, contrary to the hypotheses, raters resisted other contextual pressures that suggested that it was appropriate, even expected, for them to record inaccurate ratings. This suggests a more powerful effect of accountability than what has previously been suggested (Tetlock, 1985). The relationship between accountability and decision making may be more extensive than just decision makers complying with an ambiguous decision making environment by providing a decision that is easiest to justify. In this study, accountable raters avoided making rating decisions that were consistent with situational norms but not justifiable based on the true subordinate performance. For example, although non-accountable raters complied with the pressure to rate women higher than their performance would support, accountable raters did not do so.

Effects of Accountability on Observational Accuracy

Accountable subjects also had more accurate recall of specific performance events than their non-accountable counterparts in 2 of the 4 rating outcome conditions. This suggests that the need to justify performance ratings affects how the raters approach the rating task, a finding that was supported by the structural analysis.

It was predicted that raters who would have to account for their ratings would view each rating episode as providing evidence that will support their justification. In this case raters would, and did, carefully attend to each behavioral episode. The attentiveness resulted in accountable raters demonstrating better recall of those performance events.

Relationship Between Observational Accuracy and Differential Accuracy

Whereas observational accuracy was related to differential accuracy this study found no evidence that subjects who were trying to justify a performance rating other than the one deserved were processing performance information differently as a result. Two results support this finding. First, accountable subjects in the inflationary condition provided the most inflated (favorable) ratings of any group within the study. However, even though they complied with contextual pressure and inflated their ratings, they showed no differences in how well they recalled specific performance events. In fact, their observational accuracy was the highest in the study. Second, results of the structural analysis suggest that factors other than observational accuracy were responsible for differences in differential accuracy.

Although there were differences in how well subjects recalled specific performance events, those differences were the result of the accountability manipulations. Differences in observational accuracy did not appear to be related to the raters' desire to provide a specific performance ratings. As a result, this study found no evidence that rating outcomes affected rater information processing.

Instead it appears that raters in some conditions simply marked a rating they thought was more acceptable based on the context, not based on a skewed sample of stored behaviors.

Effects of Rating Outcomes on Rating Quality

There were no overall significant effects for the manipulations of rating outcomes. However, there were interactions effects between accountability and rating outcome for dependent variables of the overall mean ratings and the female mean ratings. While differences in rating outcomes affected rating quality, that effect was moderated by the accountability manipulation. Accountable subjects in the inflationary condition complied with situational pressures and inflated their ratings. Accountable subjects in the equitable treatment condition were reluctant to do because inflating the ratings of only female subordinates would obscure real performance differences among ratees and as a result would have been hard to justify. However, non-accountable subjects who, as predicted, were relatively resilient to contextual pressures in the other situations, significantly inflated the ratings of female subordinates when presented with a situation where it was desirable to do so.

When subjects were placed in a context where a desirable rating outcome could be achieved by rating accurately, no effect was found for either accountable or non-accountable subjects. However, it should be noted that accountable subjects in this condition were compared to other accountable subjects who had also rated more accurately than their non-accountable counterparts. For there to have been an interaction effect between the accuracy outcome and accountability conditions, accountable raters in the accuracy outcome condition would have to have been significantly more accurate than accountable raters in the comparison group. The non-significant improvement in rating accuracy found in this condition may be due to a ceiling effect for ratings accuracy. Because the design of this study provided raters multiple tasks and since performance ratings were based on several performance episodes occurring over a two week period, it is possible that there are some limitations to the extent to which rating accuracy can be improved. As a result, the inconsistent effect of rating outcome on rating accuracy may be due to unique characteristics of the dependent variable used.

These result suggest a complex relationship between competing contextual pressures within the rating environment. Raters are influenced by a variety of competing constituents who have different expectations about the performance ratings. Supervisors may prefer accuracy, subordinates may prefer favorability, and raters themselves may have reasons to prefer something else. This study found that rating behavior is the result of a complex interaction of those competing pressures. As a result, it highlights the need for research that would help us identify those pressures and determine their relationship to each other.

Accountability and Process Variables

This research also explored ways that accountability influenced rating accuracy. The structural model presented in Figure 3-4 suggests that accountability may lead to improved rating accuracy through at least three different process variables.

Attentiveness

First, accountability was strongly related to the subject's attentiveness to performance information. Subjects who were more attentive recalled specific performance information and rated their subordinates more accurately. It seems that subjects, as a result of increased attentiveness, gathered a larger, more accurate sample of performance information than raters who were less attentive. This would certainly explain the strong correlation between attentiveness and observational accuracy (*r*=.47). It is also consistent with propositions by several researchers (Motowidlo, 1986; Wherry & Bartlett, 1982) that raters who rely on a larger, more accurate, sample of behaviors may be better prepared to make rating decisions.

Note Taking

Second, accountable raters took more notes, and while they were not allowed to use their notes when making their ratings, there was a relationship between the quality and quantity of notes and the accuracy of recall of specific performance events. This finding is consistent with suggestions by others (Bernardin & Buckley, 1981; Buckley & Bernardin, 1980) that training raters to keep accurate diaries of subordinate performance should improve rating quality.

The results of the structural analysis suggest that note taking was not significantly related to rating accuracy after other variables were controlled for. Perhaps if subjects had been allowed to use their notes when they made their ratings, the relationship between observational accuracy and differential accuracy would have been significant.

Engagement

A third variable, engagement, was related to both accountability and differential accuracy. Accountable raters in this study reported they frequently debated alternative responses to the simulation and were more likely to spend time thinking about aspects of the simulation between the first and second week's session. This is consistent with suggestions by Tetlock (1985) and Schlenker

(1986) that raters will utilize more cognitively complex decision making strategies.

If the variable of engagement taps a dimension of more intensive cognitive processing of material presented in the simulation, this would explain its strong relationship to accuracy in making performance ratings. It was argued that rating subordinate performance within the dynamic environment of a typical organization is a complex task. It should be no surprise then that subjects who spent more time cognitively engaged in that task would make better rating decisions.

Summary of Process Variables

The three variables of attentiveness, engagement, and note taking provide support for the contention that the requirement to justify ratings may motivate subjects to take steps to better prepare themselves for making that rating. In this study, it appears that accountable subjects did just that by being more attentive to performance information, taking more notes on that performance information, and by spending more time thinking about situations presented in the simulation. The structural analysis discussed in Chapter 3 shows that all of the effect of accountability on both observational and differential accuracy can be explained by the three process variables.

One interesting finding from the structural analysis was that observational accuracy did not directly affect accountability. This may be due to the experimental manipulations where raters in two conditions were placed in a situation where differences in rating outcomes influenced them to make ratings they probably knew were not accurate. As discussed earlier, raters in these conditions did not have significantly lower scores on observational accuracy than raters in other conditions. The inclusion of these raters in the structural analysis might explain why observational accuracy did not affect differential accuracy when the process variables were included in the model.

Accountability and The Dilution Effect

This research also considered the possibility that accountability causes decision makers to rely on irrelevant information (Tetlock & Boettger, 1989). This research found no evidence of this "dilution effect". If accountable raters were giving weight to irrelevant information they should have been less accurate. Instead, accountable raters were more accurate. Additionally, accountable raters did not report remembering significantly more positive or negative incidents of subordinate performance than non-accountable raters.

The lack of a dilution effect may be due to the nature of the decision task. In the research reported by Tetlock and Boettger (1989) subjects made predictions about an individual based on a list of information about that target that included both relevant and irrelevant diagnostic information. The subjects were not provided any criteria to assist them in separating relevant from irrelevant information. In this study, subjects were provided detailed behaviorally anchored rating scales that clearly identified for them examples of relevant subordinate behavior. It is possible that accountability may lead to the dilution effect, only when accountable decision makers are not given adequate information about the appropriate criteria relevant to the decision task.

Implications for Future Research

Replicability

This study found that contextual factors can have a significant influence on the quality of performance ratings. The results suggest a strong and pervasive effect of accountability on the accuracy of performance ratings. Since this is the first study that used accuracy as a dependent variable, future research should attempt to replicate this study in performance appraisal or similar decision making environments.

The structural model developed from this research shows that accountable subjects responded differently to the rating task. They were more aggressive information seekers. As stated previously accountable raters were more attentive, took more notes of better quality, and were more engaged in the rating process. It seems obvious that these behaviors are relevant to improved decision quality in a wide variety of decision environments. One environment that has great similarity is the selection interview. Eder (1989) proposed that accountability be studied within the context of the interview where similar cognitive effort is required by interviewers. A study of this nature would provide an interesting environment for replication of this study.

Developing an Operational Definition of Accountability

Up to now, accountability has been treated as a dichotomous variable, subjects are either accountable or not accountable. However, within diverse organizational contexts there are probably varying levels of accountability. It is likely that there are organizational conditions which increase or decrease how accountable an individual feels about a task. In this study, subjects were told that the amount of extra credit they would receive would be based on their performance on the simulation. Accountable subjects were also told their written justification would be an important factor in that decision. This was added to the manipulation to simulate the "power" a supervisor has over an accountable subordinate. One interesting question is whether the effect of accountability would have been as strong if subjects had felt the researchers had no "power" over them? While there was no differences between the concern accountable subjects felt about getting all five points from that felt by non-accountable subjects, the power manipulation was still an important part of the overall study design.

Additionally, how are the effects of accountability different when the individual is accountable to organizational members that are not their supervisor? Klimoski and Inks (1990) found that raters who were required to justify their rating to the ratee inflated the rating as compared to raters who were told their ratings would be anonymous. In that situation (i.e. accountability to the ratee) organizational power was probably not a factor. Empirical exploration of potential levels and types of accountability is an important next step.

Expanding Models of the Rating Context

Future investigations should also consider other relevant contextual variables in addition to accountability and rating outcomes. Murphy and Cleveland's (1991) thorough review of environmental and organizational contextual influences offers many interesting directions for future laboratory and field studies that could be considered.

The Mediating Effect of Process Variables

While the section above suggests that future research should consider contextual variables that affect rating quality, it is just as important to understand how that effect occurs. While research that finds that holding raters accountable leads to more accurate ratings is interesting, it is not as interesting as research that shows what accountable raters actually do that makes them more accurate.

The structural analysis presented in this study was a first step at considering these possible influences. The model developed here should be tested on an independent sample and future research should explore additional variables that link contextual factors with rater behavior.

Summary

The results of this study support the continued exploration of the effects of contextual variables on ratings quality. In this study, accountable subjects demonstrated better recall of specific performance events and more accurately rated subordinate performance across four subordinates and three rating dimensions.

The results of this study also suggest that accountability can serve to mitigate the effects of other contextual variables which may pressure raters to provide ratings that can not be easily justified. Additionally, the data obtained from this study suggests that accountability improved performance ratings because it influenced raters to exert more effort into observing, recording, and considering the implications of subordinate performance.

This study demonstrates that raters are sensitive to a variety of organizational pressures that have a varied impact. This provides empirical support for the proposition that rating quality is strongly related to conditions of the rating context. As a result, it is consistent with the idea that rating quality may be more affected by the rater's willingness to give quality ratings than by the rater's ability (Banks & Murphy, 1985). It is time for organizational researchers to look at the entire context of the rating environment to consider whether that context influences the "willingness" of raters to provide quality ratings.

REFERENCE LIST

Abelson, R.P. (1976). Script processing in attitude formation and decision making. In J.S. Carroll & J.W. Payne (Eds.) Cognition and Social Behavior. Hillsdale NJ: Erlbaum.

Anderson, J.C., & Gerbing, D.W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. Psychological Bulletin, 103, 411-423.

Arvey, R.D., & Campion, J.E. (1982). The employment interview: A summary and review of recent research. Personnel Psychology, 35, 281-322.

Ashton, R. H. (1992). Effects of justification and a mechanical aid on judgement performance. Organizational Behavior and Human Decision Processes, 51, 416-446.

Ballaam, L.N. (1963). Multiple comparisons-A sampling experiment. Australian Journal of Statistics, 5, 62-84.

Balzer, W.K. (1986). Biases in the recording of performance-related information: The effects of initial impression and centrality of the appraisal task. Organization Behavior and Human Decision Processes, 37, 329-347.

Banks, C.G., & Murphy, K.R. (1985). Toward or narrowing the research-practice gap in performance appraisal. Personnel Psychology, 38, 335-345.

Bargh, J.A. (1984). Automatic and conscious processing of social information. In R. Wyer & T. Srull (Eds.) Handbook of Social Cognition (Vol 2). Hillsdale, NJ: Erlbaum.

Baumeister, R.F. (1982). A self-presentational view of social phenomena. Psychological Bulletin, 91, 3-26.

Becker, B.E., & Cardy, R.L. (1986). Influences of halo error on appraisal effectiveness: A conceptual and empirical reconsideration. Journal of Applied Psychology, 71, 662-671.

Bernardin, H.J. (1986). A performance appraisal system. In R.A. Berk (Ed.), Performance Assessment, (pp. 77-303). Baltimore: Johns Hopkins University Press.

Bernardin, H.J., & Beatty, R. W. (1984). Performance Appraisal: Assessing Human Behavior at Work. Belmont, CA: Wadsworth.

Borman, W.C. (1977). Consistency of rating accuracy and rater errors in the judgement of human performance. Organizational Behavior & Human Performance, 20, 238-252.

Borman, W.C. (1979). Format & training effects on rating accuracy and rater errors. Journal of Applied Psychology, 64, 410-421.

Cantor, N., & Mischel, W. (1977). Traits as prototypes: Effects on recognition memory. Journal of Personality and Social Psychology, 35, 38-48.

Cantor, N., & Mischel, W. (1979). Prototypes in person perception. In L. Berkowitz (Ed.), Advances in Experimental Social Psychology (Vol. 12). New York: Academic Press.

Cleveland, J.N., & Murphy, K.R. (1992). Analyzing performance appraisal as goal-directed behavior. In K.M. Rowland & G.R. Ferris (Eds.) Research in Personnel and Human Resource Management (Vol 10) (pp. 121-184). Greenwich, CT: JAI Press.

Cleveland, J.N., Murphy, K.R., & Williams, R.E. (1989). Multiple uses of performance appraisal: Prevalence and correlates. Journal of Applies Psychology, 74, 130-135.

Cooper, W. (1981). Ubiquitous halo. Psychological Bulletin, 90, 218-244.

Crocker, J. (1981). Judgement of covariation by social perceivers. Psychological Bulletin, 90, 272-292.

Cronbach, L.J. (1955). Processes affecting scores on understanding of others and assumed "similarity". Psychological Bulletin, 52, 177-193.

DeNisi, A.S., Cafferty, T.P., & Meglino, B.M. (1984). A Cognitive view of the appraisal process: A model research propositions. Organizational Behavior and Human Performance, 33, 360-396.

DeNisi, A.S., & Williams, K.J. (1988). Cognitive approaches to performance appraisal. In K.M. Rowland & G.R. Ferris (Eds.) Research in Personnel and Human Resource Management (Vol 6) (pp. 109-155). Greenwich, CT: JAI Press.

Driscoll, L.A., & Goodwin, W.L. (1979). The effects of varying information about the use and disposition of results on university students evaluation of faculty and courses. American Educational Research Journal, 16, 25-37.

Ebbesen, E.B., & Konecni, V.J. (1980). On the external validity of decision-making research: What do we know about decision-making in the real World? In T.S. Wallsten (Ed.), Cognitive Processes in Choice and Decision Behavior (pp.21-45). Hillsdale, NJ: Erlbaum.

Eder, R.W. (1989). Contextual effects on interview decisions. In R.W. Eder & G.R. Ferris (Eds.), The Employment Interview: Theory, Research, and Practice. Newbury Park, CA: Sage.

Eder, R.W., & Buckley, M.R. (1988). The employment interview: An interactionist perspective. In G.R. Ferris & K.M. Rowlands (Eds.), Research in Personnel and Human Resource Management (Vol 6) (pp.75-107). Greenwich, CT: Jai Press.

Endler, N.S., & Magnusson, D. (1976). Toward an interactional psychology of personality. Psychological Bulletin, 83, 956-974.

Feldman, J.M. (1981). Beyond attribution theory: Cognitive processes in performance appraisal, Journal of Applied Psychology, 66(2), 127-144.

Fisher, C.D. (1974). Transmission of positive and negative feedback to subordinates: A laboratory investigation. Journal of Applied Psychology, 64, 533-540.

Frederiksen, N., Jensen, O. & Beaton, A. (1972). Prediction of Organizational Behavior. New York: Pegamon Press.

Funder, D.C. (1987). Errors and mistakes: Evaluating the accuracy of social judgement. Psychological Bulletin, 101 (1), 75-90.

Gordon, R.A., Rozelle, R.M., & Baxter, J.C. (1988). The effect of applicant age, job level, and accountability on the evaluation of job applicants. Organizational Behavior and Human Decision Processes, 41, 20-33.

Hastie, R., & Park, B. (1986). The relationship between memory and judgement depends on whether the judgement task is memory based or on-line. Psychological Review, 93, 256-268.

Ilgen, D.R., & Feldman, J.M. (1983). Performance appraisal: A process focus. In L. Cummings & B. Staw (Eds.), Research in Organizational Behavior. (Vol 5) (pp. 141-197). Greenwich, CT: JAI

Ilgen, D.R., & Favero, J.L. (1985). Limits in generalization from psychological research to performance appraisal processes. Academy of Management Review, 10, 311-321.

Joreskog, K.G., & Sorbom, D. (1993). LISREL 8 User's Reference Guide. Chicago: Scientific Software International, Inc.

Jones, E.E. (1979). The rocky road from acts to disposition. American Psychologist, 34, 107-117.

Jones, E.E., & Nisbett, R.E. (1972). The actor and observer: Divergent perception of the causes of behavior. In E.Jones, D. Kanhouse, H. Kelley, R. Nisbett, S. Valins, & B. Weiner (Eds.). Attribution: Perceiving the Causes of Behavior. Morristown, NJ: General Learning Press.

Jones, E.E., & Wortman, C. (1977). Ingratiation: An Attributional Approach. Morristown, N.J. General Learning Press.

Kelly, H.H. (1971). Attribution in social interaction. In E.Jones, D. Kanhouse, H. Kelley, R. Nisbett, S. Valins, & B. Weiner (Eds.). Attribution: Perceiving the Causes of Behavior. Morristown, NJ: General Learning Press.

Klimoski, R., & Inks, L. (1990). Accountability forces in performance appraisal. Organization Behavior and Human Decision Processes, 48, 70-88.

Landy, F.J., & Farr, J.L. (1980). Performance rating. Psychological Bulletin, 87, 72-107.

Landy, F.J., & Farr, J.L. (1983). The Measurement of Work Performance Methods, Theory and Applications. New York: Academic Press.

Landy, F.J., Zedeck, S., & Cleveland, J. (1983). Performance Measurement and Theory. Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.

Latane, B., Williams, K., & Harkins (1979). Many hands make light the work: The causes and consequences of social loafing. Journal of Personality and Social Psychology, 37, 822-832.

Lawler, E.E. (1976). Control system in organizations. In M. Dunnette (Ed.) Handbook of Industrial and Organizational Psychology, Chicago: Rand-McNally.

Locke, (1986). Generalizing from laboratory to field: Ecological validity or abstraction of essential elements? In E.A. Locke (Ed.) Generalizing from Laboratory to Field Settings, Lexington, MA: Lexington Books

Longenecker, C.O., Sims, D.A., & Gioia, D.A. (1987).Behind the mask: The politics of employee appraisal. Academy of Management Executive, 1, 183-193.

Lord, R.G. (1985). An informational processing approach to social perceptions, leadership and behavioral measurement in organizations. Journal of Applied Psychology, 70, 66-71.

McCall, M.W., & DeVries, D.L. (1976). Appraisal in context: Clashing with organizational realities. Paper presented at the meeting of the American Psychological Association, Washington D.C.

Meier, R.A., & Feldhusen, J.F. (1979). Another look at Dr. Fox: effect of stated purpose of evaluation, lecturer expressiveness, and density of lecture content on student ratings. Journal of Educational Psychology, 71, 339-345.

Mitchell, T., (1974). Expectancy models of job satisfaction, occupational preference, and effort. Psychological Bulletin, 81, 1053-1077.

Mohrman, A.M., & Lawler, E.E. (1983). Motivation and performance appraisal behavior. In F. Landy, S. Zedeck, & J. Cleveland (Eds.). Performance Measurement and Theory. Hillsdale, NJ: Erlbaum.

Motowidlo, S.J. (1986). Information processing in personnel decisions. In K.M. Rowland & G. Ferris (Eds.), Research in Personnel and Human Resource Management, (Vol 4) (pp. 1-44).

Murphy, K.R., & Balzer, W.K. (1986). Systematic distortions in memory-based behavior ratings and performance evaluations: Consequences for rating accuracy. Journal of Applied Psychology, 71, 39-44.

Murphy, K.R., Balzer, W.K., Kellam, K.L., & Armstrong, J.G. (1984). Effects of the purpose of rating on accuracy in Observing teacher behavior and evaluating teaching performance. Journal of Educational Psychology, 76(1), 45-54.

Murphy, K.R., & Cleveland, J.N. (1991). Performance appraisal an organizational perspective. Boston: Allyn and Bacon.

Murphy, K.R., Garcia, M., Kerkar, S., Martin, C., & Balzer, W.K. (1982), Relationship between observational Accuracy and Accuracy in Evaluation Performance. Journal of Applied Psychology, 67(3), 320-325.

Murphy, K.R., Philbin, T.A., & Adams, S.R. (1989). Effect of purpose of observation on accuracy of immediate and delayed performance ratings. Organizational Behavior and Human Decision Processes, 43, 336-354.

Porter, L.W., & Lawler, E.E. (1968). Managerial Attitudes and Performance. Homewood, IL: Dorsey Press.

Saal, F.E., & Knight, P.A. (1988). Industrial/Organizational Psychology: Science and Practice. Pacific Grove, CA: Brooks/Cole.

Schlenker, B.R. (1980). Impression Management: The Self Concept, Social Identity, and Interpersonal Relations. Monterey, CA: Brooks-Cole.

Schlenker, B.R. (1982). Translating actions into attitudes: An identity-analytic approach to the explanation of social conduct. In L. Berkowitz (Ed.), Advances in Experimental Social Psychology (Vol 15). New York: Academic press.

Schlenker, B.R. (1986). Personal Accountability: Challenges and Impediments In the Quest for Excellence. (Contract No. 400-77-0099). Navy Personnel Research and Development Center.

Schwab, D.P., Olian-Gottlieb, J., & Henneman, H.H. (1979). Between subjects expectancy theory research: A statistical review of studies predicting effort and performance. Psychological Bulletin, 86, 139-147.

Sheffrin, R.M., & Schneider, W. (1977). Controlled and automatic human information processing II: Perceptual learning, automatic attending, and a general theory. Psychological Review, 84, 127-190.

Shepperd, J.A. (1993). Productivity loss in performance groups: A motivation analysis. Psychological Bulletin.

Simonson, I., & Nye, P. (1992). The effect of accountability on susceptibility to decision errors. Organizational Behavior and Human Decision Processes, 51, 416-446.

Srull, T.K., & Wyer, R.S. Jr. (1979). The role of category accessibility in the interpretation of information about persons: Some determinants and implications. Journal of Personality and Social Psychology, 37, 1660-1672.

Sulsky, L.M., & Blazer, W.K. (1988). Meaning and measurement of performance rating accuracy: Some methodological and theoretical concerns. Journal of Applied Psychology, 73(3), 497-506.

Taylor, S.E., & Fiske, S.T. (1978). Salience, attention, and attributions: Top of the Head Phenomena. In L. Berkowitz (Ed.) Advances in Experimental Psychology (Vol 11) (pp. 249-288). New York: Academic Press.

Taylor, E., & Wherry, R. (1951). A study of leniency of two rating systems. Personnel Psychology, 4, 39-47.

Tetlock, P.E. (1983). Accountability and complexity of thought. Journal of Personality and Social Psychology, 45, 74-83.

Tetlock, P.E. (1985). Accountability: The neglected social context of judgement and choice. In B.M. Staw & L. Cummings (Eds.), Research in Organizational Behavior (Vol 7) (pp. 297-332). Greenwich, CT: JAI Press.

Tetlock, P.E., & Boettger, R. (1989). Accountability: A social magnifier of the dilution effect. Journal of Personality and Social Psychology, 57, 388-396.

Tetlock, P.E., & Kim, J.I. (1987). Accountability and judgement processes in a personality prediction task. Journal of Personal and Social Psychology, 52, 700-709.

Tetlock, P.E., Skitka, L., & Boettger, R. (1989). Social and cognitive strategies for coping with accountability: Conformity, complexity and bolstering. Journal of Personality and Social Psychology, 57, 632-640.

Vroom, V. (1964). Work and motivation. New York: Wiley

Waldman, D.A., & Thornton, G.C. III (1988). A field study of rating conditions and leniency in performance appraisal. Psychological Reports, 63, 835-840.

Weiner, (1972). Theories of Motivation: From Mechanism to Cognition. Chicago: Markham.

Wherry R.J. (1952). The Control of Bias in Rating: A Theory of Rating. PRB Report No. 922. Contract No. DA-49-083 05A 69, Department of the Army.

Wherry, R.J. Sr., & Bartlett, C.J. (1982). The control of bias in ratings: A theory of ratings. Personnel Psychology, 35, 521-551.

Williams, K.J. (1984). Appraisal Salience and the Need to Reprocess Information: Effects on Organization, Recall, and Rating of Performance Information. Ph.D. dissertation. Department of Psychology, University of South Carolina.

Williams, K.J., DeNisi, A.S., Blencoe, A.G., & Cafferty, T.P. (1985). The role of appraisal purpose: Effects of purpose on information acquisition and utilization. Organizational Behavior and Human Performance, 35, 314-339.

Williams, K.D., Harkins, S.G., & Latane, B. (1981). Identifiability as a deferent to social loafing: Two cheering experiments. Journal of Personality and Social Psychology, 40, 303-311.

Wortman, C.B., & Linsenmeier, J.A. (1977). Interpersonal attraction and techniques of magnification in organizational settings. In B.M. Staw & G. Salancik (eds.). New Dimensions in Organizational Behavior. Chicago, IL: St Clair's Press.

Zedeck, S., & Cascio, W.F. (1982). Performance appraisal decisions as a function of rater training and purpose of the appraisal. Journal of Applied Psychology, 67, 752-758.

APPENDIX A
SAMPLE IN-BASKET ITEMS

THE CONSOLIDATED FEDERAL AGENCY CONTRACTING OFFICE

A MANAGERIAL IN-BASKET SIMULATION

Consolidated Federal Agency Contracting Office First In-Basket Exercise

Situation

It is 7:45 A.M. on Wednesday, May 20, you have recently been placed in charge of the division due to the unexpected departure of your predecessor, Bob Ward. Your selection to this position occurred very quickly and unfortunately, you do not know the organization as well as you would like. You have worked as a supervisor in the purchasing department for a large business organization but this is your first experience in government service. You have a busy schedule planned for the day and you have decided to allocate any time available between appointments to familiarize yourself with the organization and to read and respond to the large number of administrative items that your assistant has placed in your in-basket. Your schedule calls for you to meet with your boss Art Bollinger at 4:30 so he can introduce you to several directors of the federal agencies that you will be working with. You are also scheduled to leave this evening at 6:00 P.M. for a trip to Chicago with Art and Wanda Lane, the Chief of the Human Resource Division. As a result of this schedule, it is important that you accomplish as many tasks in your in-basket as possible. You have also made it known to your staff that you will be in your office this morning and that they should feel free to drop by so you can get to know them better.

Instructions

You should go through the papers in your in-basket in any way you wish and take whatever action you think appropriate. The television monitor is an integral part of your work environment and it will allow you to observe first hand the situations that occur in your organization. It is important that you remember that the camera recording these incidents is your character, Leslie Wilder, and as a result, you will observe what Leslie sees and hears over the course of a normal workday. The information provided during the situations you will see will be important for decisions you will need to make on organizational problems. While you can't respond directly to the situations you will witness, it is important that you write down any actions you would have taken. Make sure you write down every action that you take in response to the in-basket or video inputs. Write your responses directly on the in-basket items using margins or the back of the forms as desired. Include notes to your subordinates, yourself, and others. Keep in mind that your ideas and actions can't be evaluated unless you write them down. If you should need additional space, blank paper is provided at the end of the in-basket, Do not just say what you would do in the situation, but actually write letters, memos, etc. where appropriate.

An organization chart, a six week calendar, and a daily calendar have been provided to assist you in decision making and to help you maintain your schedule.

CONSOLIDATED FEDERAL AGENCY
CONTRACTING OFFICE

Memo

To: All Management Level Staff
From: Wendy
Date: May 18
Subject: Staff Meeting

Leslie has asked me to distribute a reminder that we will have a staff meeting for management level personnel on Tuesday, 26 May. The meeting will begin at 2:00 P.M. in the conference room.

To help Leslie become familiar with the areas of importance within each section, please prepare a presentation covering the following topics:

- Goals for your unit for the next twelve months
- Significant accomplishments
- Potential problem areas
- Staffing status
- Special interest items

Also, please provide copies of your presentation information.

See you on Tuesday!

Wendy

cc: Leslie Wilder

CFACO IS AN EQUAL OPPORTUNITY EMPLOYER

CONSOLIDATED FEDERAL AGENCY
CONTRACTING OFFICE

Memo

To: Leslie Wilder
From: Art Bollinger
Date: 18 May
Subject: Lunch

Leslie,

Welcome aboard! I am sure you'll find your responsibilities challenging and exciting. I am looking forward to our trip to attend the Federal Employment Regulation Conference since it will be a good chance for us to get to know each other better. I also would like for you to leave some space in your schedule around 4:30 on the 20th. I will be concluding a meeting with several directors of the federal agencies that we service and I would like you to come down to meet them. I'll have my secretary call when we finish up.

Also, let me know if you can have lunch with me next Wednesday at 12:30 P.M.

Art

CONSOLIDATED FEDERAL AGENCY
CONTRACTING OFFICE

Memo

To: Bob Ward
From: Human Resources Training Division
Date: 13 May
Subject: Annual Government Contracting Training Course

We have received the schedule for this years Government Contracting Training Course. This training is mandatory for all government agents who are responsible for approving contracts in excess of $10,000. I believe this includes all members of your staff. This is a three day course and is offered on Tuesday, Wednesday and Thursday of the first three weeks of July.

Please provide me a list of who will be attending this course and their desired dates. I will do my best to get their first choice of class dates.

Miles Sloan
Chief of Training, Human Resources Division

CFACO IS AN EQUAL OPPORTUNITY EMPLOYER

CONSOLIDATED FEDERAL AGENCY
CONTRACTING OFFICE

Memo

To: Leslie Wilder
From: Steve Flannery
Date: May 18
Subject: Lunch

Welcome aboard Leslie. I was hoping you would have some time on your schedule next Wednesday or Thursday, May 27th or 28th, for lunch. I think it would be a great time to get acquainted since we will be working together very closely in the future. Let me know how that fits with your schedule!

Steve

CFACO IS AN EQUAL OPPORTUNITY EMPLOYER

CONSOLIDATED FEDERAL AGENCY
CONTRACTING OFFICE

Memo

To: All Supervisors
From: Wanda Lane
Date: May 19
Subject: Annual Position Review

You will find attached a final copy of job descriptions for supervisory personnel within your division. Thank you and your staff for involvement in this review. I can't stress how important it is for our organization to maintain an accurate job description for each position. This document is critical to my division providing you qualified applicants when you have a job opening. It is also important in maintaining a fair and competitive compensation program for our employees.

Thanks for your participation in this review. Please notify my office if anything changes relative to your position descriptions prior to our next annual review.

Wanda L. Lane
Director of Human Resources

Atch 1
Annual position
review

CFACO IS AN EQUAL OPPORTUNITY EMPLOYER

Job Title: Regional Contracting Manager
ID Number: 07687-511-3321
Division: Administrative and MIS Contracting

Job Summary
Supervising all contracting activities in support of stated customers needs.

Job Duties
Managing a staff responsible for:

1. Reviewing customers stated requirements and related specifications and insuring all documents are properly recorded and are in compliance with applicable Federal, State and Local directives.

2. Preparing and distributing requests to suppliers for bids on routine and non-routine supplies.

3. Collecting and evaluating supplier bids based on stated customer specifications, cost and other factors.

4. Negotiating contract rates and specifications. Resolving conflicts between customers and suppliers.

5. Monitoring contract compliance, conducting routine follow-up inspection on quality of goods and services. Monitoring customer satisfaction and contractor compliance with applicable regulations.

Skills, Knowledge, Abilities

Education Needed: Bachelors Degree in Contracting, Logistics. Business or a related field.

Training Needed: Federal Supervisors Orientation Course
Initial Government Contractor Training Course
Recurring Government Contractor Training

Machine, Tools Equipment, Work Aids
None

Relationship To Other Jobs
Promotion From: Contracting Support Specialist

Promotion to: Chief of Contracting

Supervision Received: Chief of Contracting

Supervision Given: Contracting Support Specialist
Administrative Assistant/Secretarial Support

Current Staff In Job
Edward Montage
Carole Donaldson
David Fredericks

CONSOLIDATED FEDERAL AGENCY
CONTRACTING OFFICE

Memo

To: Leslie Wilder
From: Alice Garcia
Date: 15 May
Subject: Staffing Policy

I have been working with the Human Resource Department to hire a new secretary/assistant. I have narrowed the choice down to three candidates and I would like to make a final decision as soon as possible because it has been very difficult to operate without an assistant.

I would like to know how extensive an involvement you would like to have in our hiring decisions. I have included a copy of Bob Ward's policy on internal staffing procedures -- will yours be different? I am planning on working on this while you are in Chicago and I have set up interviews for Friday morning so I can resolve this before my trip to Seattle. If I can get a draft of your policies on this issue it would help matters a great deal. Thanks for any guidance you can offer. By the way, welcome to the organization!

Alice Garcia
Contract Compliance Manager

Atch 1
Procurement
Hiring Procedures

CFACO POLICY NOTICE

To: All Staff
From: Bob Ward, Chief of Administrative and MIS Contracting
Subject: Procurement Division Staffing Procedures

Staffing is one of the most important activities we are charged with. The success of our organization depends on the quality of personnel we hire. I am very interested in taking an active role in the selection of each member of our division. I am therefore implementing the following policies:

1. Each manager will contact the Human Resource Division when a job opening in their division occurs.

2. Individual managers will be responsible for insuring that adequate job requirements are correctly outlined in the applicable job description.

3. Individual managers will monitor the application and screening process to insure our interests are being well represented by the HR staff.

4. Individual managers may review records, applications, resumes etc. as appropriate to assist in the applicant screening process.

5. Individual managers will select three applicants for final consideration and interview by the Chief of Administrative and MIS Contracting. Application packages are forwarded to the Chief 72 hours prior to the interview.

6. At the conclusion of the interviews, the Chief of Administrative and MIS Division, in consultation with the appropriate manager, will make the hiring decision.

7. Individual managers will be responsible for notifying the HR department and unsuccessful applicants of the final decision.

I think you will agree that this process will help us select the very best employees possible.

Bob Ward

May 15

Big Brother's of America
836 N. 3rd Street

Leslie Wilder
Chief Administrative and MIS Division
CFACO
1123 N. Manhattan

I want to express my sincere thanks to John Shroud of your staff who has devoted countless hours to supporting the Big Brother program. Not only is John an excellent big brother himself, but he recently volunteered to head up our annual fund raising effort to send underprivileged kids to camp for the summer. As a direct result of John's tireless efforts, we will be able to send 50 of these boys to camp this summer. This is by far the largest total we have ever supported and we have John to thank for that.

It must be a real pleasure to have someone of John's calibre on your staff. I'm sure you will agree with me that he is an excellent individual. Please pass on not only my personal thanks, but the thanks of all the children he has helped through his work with the Big brothers program.

Alfred Hoight
Chapter President

Consolidated Federal Agency Contracting Office Second In-Basket Exercise

SCENARIO

Situation

It is now Monday, June 8 at 1:20 P.M. At the beginning of this exercise, you and your assistant Wendy are in the division conference room where a meeting with Bill Jensen and Ed Montage to discuss their section's travel budgets is just finishing. In addition to attending a division staff meeting at 2:00 P.M., you are going to try to respond to the material in your in-basket since it has be almost two weeks since the last time you have been able to work through it.

You are still dealing with many of the issues that came to your attention during the first in-basket. By way of reminder, here are the issues you were dealing with:

1. An effort was underway by Steve Flannery, the Chief of Services Division, to move responsibility for the MIS effort to his division.
2. Janet Keen, who works in your Central Region was concerned about her relationship with her supervisor and your subordinate Ed Montage.
3. Retention and replacement of contracting specialists appeared to be a problem and CFACO is currently under a federal hiring freeze.
4. Travel expenses for CFACO as a whole had exceeded the budget and more restrictions on travel were required.

Attached are a copies of key excerpts from you notebook where you keep all notes, reminders, and ideas. These notes will provide you information about meetings and discussions that have occurred in the two-week period since you were last able to work through your in-basket. To assist you, we have included copies of your schedules for the period as well as the CFACO organization charts.

Instructions

You are to go through the papers in your in-basket in any way you wish and take whatever action you think appropriate. The television monitor is an integral part of your work environment and it will allow you to observe first hand the situations that will occur in your organization. It is important that you remember that the camera recording these incidents is your character, Leslie Wilder, and as a result, you will observe what Leslie sees and hears over the course of a normal workday. The information provided during the situations you will see will be important for decisions you will need to make on organizational problems. While you can't respond directly to the situations you will witness, it is important that you write down any actions you would have taken. Make sure you write down every action that you take in response to the in-basket or video inputs. Write your responses directly on the in-basket items using margins or the back of the forms as desired. Include notes to your subordinates, yourself, and others. Keep in mind that your ideas and actions can't be evaluated unless you write them down. If you should need additional space, blank paper is provided at the end of the in-basket, Do not just say what you would do in the situation, but actually write letters, memos, etc. where appropriate.

An organization chart, a six week calendar, and a daily calendar have been provided to assist you in decision making and to help you maintain your schedule.

****

NOTEBOOK EXCERPTS

****

Flight to Chicago W/Art Bollinger and Wanda Lane May 20

- Art stressed he sees important job challenges as
  1) Controlling costs
  2) Reducing dependency on foreign suppliers and manufacturers
  3) More contact with the customer.

He would like to see us make at least one trip a month to visit some of our customers so we can "see how we are doing"

- Brought up his feelings that after a discussion with Flannery, he believes responsibility for MIS support should be transferred to Services division. He consented to allowing me to put together a position paper on the pros and cons of this idea.

NOTE: I tasked B. Jensen to work on paper and get back to me as soon as possible.

- Discussed new management development program. Current program (Wanda said she would send a copy to me this week), calls for division level managers to offer opportunities for interested subordinates to participate in a program where they "shadow" senior managers for a week or so to learn the roles and responsibilities of each position. The program also encourages Division Chiefs to rotate responsibility for filling in for them when they are gone. They also encourage division managers to report on the effectiveness of subordinates when "filling in".

- Bollinger gave me responsibility for totally revising the program. He wanted this to be one of my top priorities since the agency was currently finalizing all training requirements and he was interested in pursuing more formal training opportunities for our staff if we determined that was an "appropriate" step to take. He outlined three areas of improvement for the program

1. Increase program formality. He felt too much of the program was voluntary and that division chiefs and the HR division should take a more formal and aggressive approach to the program.
2. Create more opportunities for lower level employees to actively demonstrate and develop their management skills rather.
3. Develop more formal training opportunities including setting up programs with local colleges and universities.

- We set up a meeting on Tuesday, 9 June at 0800.

NOTE: I sent a memo to the staff to see if anyone was

interested in participating in the Management Development program and to gather inputs and ideas about the program.

Meeting with J. Keen 2 June
- Ms Keen was obviously quite distraught and came in to see me unannounced. She was following up on her Memo 11 May.
- Following is a summary of her concerns
1. She has been "passed over" for promotion opportunities and merit pay increases because of her conflict with Montage.
2. Their differences, in her opinion, revolve around her "deep religious beliefs"
3. She claims Montage made her remove religious pictures from her work area because it conflicted with a professional office image
4. She claims that profane language in the work area is permitted
5. Friends have told her Montage calls her "Miss Goody Two Shoes" behind her back

- She said she has been looking for another job but people are reluctant to consider her seriously because she won't provide her current employer as a reference

- She has been to see the Equal Opportunity and Treatment Officer but is not sure they will do anything to help her.

- I told her I was looking into the situation and asked that she give me one week to get back to her. She was still very upset when she left my office.

Meeting W/Ed Montage 26 May
- We reviewed Janet Keen's record and discussed the validity of her claims. He felt that while she had a great deal of promise she had created an environment for herself where she was disliked by her coworkers and that the situation was really her own fault. He noted that her performance ratings had diminished over the past couple of years and that he did not feel that her loss would be too great. I pointed out that while she may have received lower ratings the past two years, her ratings were still in the good to excellent range. He responded that was just "the inflated system."

- I asked about the conflict over the pictures that she claimed he asked her to remove. He stated that she had several religious statues and pictures that he felt didn't contribute to a professional image in the office. I asked if he knew if there was a written policy on the matter and he was unaware of such a policy.
- Overall, he was concerned that the matter had gotten out of his control and asked that he be allowed to handle it himself. I told him that I was committed to responding to Janet but I felt it was important that he speak with her to attempt to

resolve differences.

NOTE: I asked Wendy to find out what policies exist for personal decorating (i.e. pictures) of work areas.

Meeting With J. Keen 27 May

- I called Ms Keen in after my discussion with Montage

- I stressed the fact that I was very concerned that she felt she was being unfairly treated. I also told her I strongly supported the policies of the organization that prohibited any type of discrimination and that Montage understood my position on these issues.

- I pointed out that I was researching the company policy on personal displays in the work area and I would insure that we were complying with those policies. I also added that if there was some disagreement with those policies we would insure they were fairly reviewed to see if they were appropriate.

- Ms Keen stated she felt unable to work in the situation any longer and that she wanted to officially request a transfer. If a transfer was not granted, she was planning on leaving the organization.

- She felt that despite my efforts she would still not be able to maintain a working relationship with Montage and it was time for her to move on. If I didn't move her to a new division, she was going to leave CFACO. I asked her to be patient with the process and to allow me a few more days to talk to her coworkers to sort this matter out. She wouldn't make any promises but said she would tell me before she made any decision.

Staff Meeting Notes 26 May (Admin & MIS Division Staff)

- Since this was my first meeting with my staff I told them how much I was looking forward to working with them and in this organization. I asked that they each try to keep in mind that I am new to the federal service so they may have to provide me more background on problems and issues than would normally be required.

- Each staff member then made a presentation about their section goals, accomplishments and challenges.

- Following are questions asked of each presenter

A.Garcia

1. What are the regional breakdowns for percentages of

contracts issued for minority owned, small businesses and foreign suppliers.
2. Are their regional peculiarities which would lead us to believe some regions would have more difficulty finding minority owned or small business suppliers than other regions
3. What options do we have for your division reviewing contracts that are assigned by other divisions.

NOTE: There was an obvious concern among the regional managers that the Contract Compliance Division had too much control over each region's business. Hopefully, we can find a solution to this that will allow Alice to do her job while not making the others feel like she is constantly looking over their shoulder!

C. Donaldson

- Provide me all data supporting her contention that regulations on small and minority owned businesses are causing overall contracting costs to escalate.

NOTE: I told her I would support a presentation to Mr Bollinger if the data supports her contention!

B. Jensen

- We discussed the current movement to transfer responsibility for MIS follow-on support to Services. Bill was quite vocal in his feelings that a transfer was not appropriate. We discussed the possible reasons for and against a transfer. I told Bill that completing the installation of the MIS system should be his first priority and assured him that no decision would be made on this issue until he had a chance to make his case on reorganization.

- We also discussed client reactions to the initial turmoil created by the installation of the MIS System. There was a sense of concern from all participants.

NOTE: After discussion I asked Bill to send a letter to all CFACO clients announcing that the system would soon be operational. I stressed the need to make the letter positive and one that encouraged clients to use the system. We also discussed the need to thank those clients who had problems with the system during the installation and checkout phase.

Lunch with Steve Flannery 28 May

- Discussed the merits of the idea to move responsibility for MIS follow on service and support to his department.

- His main point is that service is "his business" and since we are no longer buying the hardware, responsibility for MIS should belong to those who have specific expertise in warranty type follow on service (e.g. his division.)

- I pointed out the lack of expertise that his division had with this type of contract and it became clear that he intended to try to get both the responsibility and the personnel in this area (Bill Jensen and staff) transferred to his department.

- I told him we were putting together information on this issue for Bollinger and that we would let him know when we were discussing the issue further. He didn't seem pleased!

CONSOLIDATED FEDERAL AGENCY CONTRACTING OFFICE

Memo

To: Leslie Wilder
From: Wendy Martin
Date: June 3
Subject: CFACO Policies on Decoration of Work Area

At your request, I checked out the policies on personal effects in the office environment with the folks in HR. They say the subject has come up before a few years ago when one employee objected to a picture from the Sports Illustrated Swimsuit edition being hung in the office. The policy states that "employees should take care to not display any material which is offensive in nature. Prohibited material includes items that have explicit sexual content, profane or suggestive language or material that is discriminatory in nature. Supervisors are responsible for insuring that personal items do not offend employees or guests and are conducive to a professional office environment."

I Hope this helps.

Wendy

CFACO IS AN EQUAL OPPORTUNITY EMPLOYER

CONSOLIDATED FEDERAL AGENCY CONTRACTING OFFICE

Memo

To: Leslie Wilder
From: Robert Deleon
Date: June 1
Subject: Washington D.C. Trip

I am scheduled to go to D.C. for three days from 22-24 June to testify before congressional staffers for the House Appropriations Committee on some financial and contractual reporting issues that affect CFACO. Art suggested that you should accompany me so you can gain experience in this area. I think it is a good idea if you can arrange it with your schedule. If you can go, call my secretary and she will set it up. Call me if you have any questions.

Bob

CONSOLIDATED FEDERAL AGENCY
CONTRACTING OFFICE

Memo

To: All Supervisors
From: Wanda Lane
Date: May 29
Subject: White House Fellows Program

Applications for this year's White House Fellows program are due to my office by June 5. This extremely competitive program selects highly qualified federal employees to participate in a one year program working with members of the White House staff. Interested applicants may review information about this excellent program by contacting Bill Lainge in the Training Section.

Supervisors should notify all eligible applicants of this opportunity and insure that those interested submit their packages by the deadline. CFACO can nominate one person for this program and that selection must be made by Monday, June 8. Good luck to all applicants.

Wanda

CFACO IS AN EQUAL OPPORTUNITY EMPLOYER

Bureau Of Banks
1011 South Orange
Buffalo, NY

28 May

Leslie Wilder
Chief Administrative and MIS Division
Consolidated Federal Contracting Office
Chicago, IL

Dear Leslie,

I have had some recent dealings with your Eastern Region staff that I feel requires your immediate attention. They had negotiated a five year contract with a supplier for printing our forms and other related paperwork. While I was very pleased with the support we received throughout the negotiation process, I must admit some recent events have changed my opinion.

We were supposed to receive our first shipment of forms from the supplier by May 1. When we had not received them by May 7, I spoke with your staff and was told they would look into the matter. My call was returned on May 9 and I was told the forms had been printed and that while there was a delay in shipment, we would have them by the 15th. We still had not received the forms by May 20 and so I personally called the supplier to find out the status of our order. I was told that the supplier had been unable to find printing equipment that met the stringent handling requirements needed to prepare our forms. The supplier thought they were close to procuring the needed equipment, but would not be able to deliver the forms until June 30. Unfortunately, our current supply of forms will be exhausted by June 15. I told your staff that this situation was unacceptable and asked them what we were supposed to do. It was clear to me that they felt this matter was our problem and not theirs and that they were unconcerned with the delay. At this point, I am afraid we will be unable to provide the required support to our member banks. I am very frustrated with this situation and I feel CFACO has done nothing to assist us. I would appreciate your immediate response to this situation.

Sincerely,

Michael P. Schmidt
Director

June 2

Dr. John Bornen
Director of Purchasing
Benton Industries
Indianapolis, IN

Mr Leslie Wilder
CFACO
Chief Administrative and MIS Division
1123 N. Manhattan

Dear Leslie,

It is my pleasure to recommend Ms. Laurie Churn for a position in your organization. Laurie has been employed in our procurement division for the past five years. In that time I have found her to be bright, energetic, articulate, and resourceful. Her innovative style has helped us refine our purchasing procedures to improve the overall efficiency and effectiveness of our purchasing operations. Over that period Laurie has been promoted from the entry position of purchasing agent, to the position of executive assistant to the director of purchasing. In this capacity, she has assisted in the day to day management of the purchasing activity for an organization that procures over 250 million dollars annually in good and services. Laurie, knows the business and has the skills and energy to use that knowledge in a way that is very effective.

Unfortunately for Benton Industries, Laurie has decided to move to your area to be with her husband. Fortunately for you, this highly capable person is available to work in your organization. I have suggested to Laurie that she contact you during her next visit to your area. I am sure after meeting her you will want to find a position in your organization for her. Please feel free to contact me if there is any additional information I can provide you about Laurie. As much as I hate to lose her, I am sure she will use her abilities for the success of some other organization.

John Bornen
Director of Purchasing

PHONE MESSAGE

DATE: 4 June

CALLER: Robert Deleon

TIME: 9:30

MESSAGE: He would like to have lunch sometime during week of 15 June. Please call and tell him what date would be good for you.

CONSOLIDATED FEDERAL AGENCY CONTRACTING OFFICE

Memo

To: All Division Chiefs
From: Wanda Lane
Date: June 5
Subject: Inputs on CFACO Turnover Discussion

I know many of you are considering your response to Art's memo (June 3, Subj: "Turnover of Contracting Specialists"). To assist you, I have included a summary of the types of responses we have been getting during exit interviews with staff who are leaving CFACO. Good luck as you wrestle with this problem.

Wanda

CFACO IS AN EQUAL OPPORTUNITY EMPLOYER

RESULTS OF EXIT INTERVIEWS WITH STAFF WHO VOLUNTARILY LEAVE CFACO

Main Reason Cited For Leaving CFACO:

22%	Lack of Advancement Opportunity
18%	Low Pay
11%	Low or Inadequate Benefits
11%	Dissatisfaction With Supervision
9%	Dissatisfaction With Duties
9%	Better Opportunity Elsewhere
7%	Other Reason or No Response
4%	Transfer of Spouse
4%	Too Much Travel
3%	To Pursue Higher Education
2%	Health Reasons

Other Reasons for Leaving:

(Note: These are open ended responses that we have attempted to summarize below. Respondents are allowed to list more than one reason)

48%	Lack of Advancement Opportunity
38%	Low Pay
33%	Low or Inadequate Benefits
20%	Dissatisfaction With Supervision
28%	Better Opportunity Elsewhere
23%	Lack of Training Opportunities
12%	Dissatisfaction With Duties
11%	Too Much Travel
9%	To Pursue Higher Education
7%	CFACO Was Too Far From Home
2%	Health Reasons
1%	Transfer of Spouse

Benefits Desired By Leavers

For those who cited low benefits as a reason for leaving, I have listed the benefits they were most concerned with:

36%	Health Care
21%	Dependent Care
11%	Annual Leave
7%	Retirement
5%	Compensation For Lost Work Time (This includes any reference to supplements paid to workers who lose their earnings due to illness or injury)

APPENDIX B
BEHAVIORALLY ANCHORED RATING SCALES

INSTRUCTIONS FOR USE OF THESE SCALES

These scales will provide you with specific information about the performance dimensions you will rate. Please read this booklet carefully and insure that you understand the behaviors associated with each rating dimension before you proceed.

You will rate your subordinates on the following performance dimensions:

Administration:

Attending to detail in subordinates' administrative work products; writing clearly and effectively; meeting timelines for completion of assignments.

Coordination and Negotiation:

Developing and maintaining good working relationships; handling conflict; dealing with stressful interpersonal situations; demonstrating receptivity to inputs and suggestions.

Supervision:

Directing, motivating, developing, and informing subordinates.

Work Effort:

Volunteering for challenging assignments; sacrificing personal convenience for task accomplishment; handling tasks beyond normal area of responsibility; handling crises.

On the following pages you will be provided a series of rating scales. Each page includes a specific performance dimension. Notice that the performance category is defined at the top of the page. Running down the left side of the page is a 7-point scale broken into three ranges with 7 and 6 at the HIGH range, 5, 4, and 3 at the MODERATE range, and 2 and 1 at the LOW range. Each range is defined with a description of what a person would have to do to deserve a rating in that range of HIGH, MODERATE OR LOW.

ADMINISTRATION

Attending to detail in subordinates' administrative work products; writing clearly and effectively; meeting timelines for completion of assignments.

7 6	HIGH	Regulates subordinate administrative work so that unit products are accurate and reflect professional detail Writes correspondence that is appropriate, clear, and concise Responds promptly and accurately to internal and external requests and timelines for work assignments
5 4 3	MODERATE	Oversees subordinate administrative work to insure normal standards are maintained Writes with appropriate format, grammar and tone Meets or arranges for an extension to internal and external requests and timelines for work assignments
2 1	LOW	Fails to insure subordinate work products meet acceptable standards Writes correspondence that is frequently vague and inaccurate and requires extensive rewrite Ignores externally generated timelines and requirements

WORK EFFORT

Volunteering for challenging assignments; sacrificing personal convenience for task accomplishment; handling tasks beyond normal area of responsibility; handling crises.

7 6	HIGH	Actively seeks and volunteers for challenging assignments Sacrifices other personal activities when necessary to attain goals Volunteers for tasks even when not strictly in own area of responsibility Takes immediate, independent, and quick action to meet a crisis or a need
5 4 3	MODERATE	Accepts challenging assignments Rearranges personal commitments if essential to attain goals Accepts assignments even when not strictly in own area of responsibility Responds quickly, and with little guidance to meet a crisis or a need
2 1	LOW	Avoids or declines assignments which are difficult or challenging Resists involvement in activities which conflict with personal interests or pursuits Refuses to take on problems or tasks when not strictly in own area of responsibility Responds impulsively or only when directed to meet a crisis or need

SUPERVISION

Directing, motivating, developing, and informing subordinates.

7 6	HIGH	Coordinates subordinate work efforts to insure accurate and efficient accomplishment of tasks Creates work environment that recognizes and rewards subordinate goal accomplishment and shows sensitivity to subordinate personal needs Develops subordinates by recognizing their strengths and correcting their weaknesses Keeps subordinates abreast of relevant information in a timely and appropriate manner
5 4 3	MODERATE	Maintains subordinate work activities towards task accomplishment Attempts to motivate others by providing feedback and encouragement and being aware of personal needs Recognizes and points out to subordinates their areas of strengths and weaknesses Informs subordinates of major events that impact the work environment
2 1	LOW	Disrupts work activities of others and thereby interferes with task accomplishment Does not motivate or encourage others to exert more effort toward task accomplishment or recognize personal needs Unaware of differences in subordinate work performance or overreacts to subordinate weaknesses and fails to recognize strengths Fails to keep subordinates abreast of relevant information

COORDINATION AND NEGOTIATION

Developing and maintaining good working relationships; handling conflict; dealing with stressful interpersonal situations; demonstrating receptivity to inputs and suggestions.

7 6	HIGH	Establishes and nurtures a positive and open communication environment Mediates conflicts through tactful and persuasive intervention Maintains composure and responds calmly to stressful situations Actively seeks the inputs and suggestions of others
5 4 3	MODERATE	Responds quickly to damaged working relationships Encourages others to consider alternative responses to conflict situations Avoids strong emotional reactions to stressful situations Accepts inputs and suggestions of others
2 1	LOW	Allows adversarial or noncooperative working relationships to exist Jumps to conclusions during conflict situations which complicates stressful situations Loses self-control in stressful situations Avoids or rejects inputs and suggestions of others

APPENDIX C
EXPERT JUDGES' INSTRUCTIONS

Dear Friends and Colleagues,

Thank you for your willingness to serve as expert raters for this project. Because your role is very important for its success, I have prepared this set of guidelines in some detail to help you complete your ratings as efficiently as possible.

Project Overview

You are going to view videotaped presentations that we are using as part of our research on performance appraisals. When we use this material for research purposes our subjects will assume the role of a manager, Leslie Wilder, in a federal organization. They will work on a series of simulated administrative tasks that typical managers might find in their "in-basket". These tasks include responding to memos, phones calls, and letters that will be used to present a variety of organizational problems that they will have to deal with through effective decision making, delegation etc. While working on the in-basket, they will be interrupted periodically by short videotaped scenes that will show their subordinates presenting information related to the in-basket. Each of those scenes was created especially to communicate something about the level of performance for that subordinate. That is where you, as an expert rater, come in. I need your best estimates of the level of performance that is being represented in each scene.

Role of Expert Raters

I would like your assistance as an expert rater for the set of videotaped scenes we will show in this simulation. In your capacity as the expert, I need you to evaluate the performance of the individuals that you will see in each scene. I will use your ratings to help me understand exactly what the behavior of each subordinate is communicating about their performance. I intended each scene to very clearly communicate information about the subordinate. Your job is to provide me feedback on what you think each scene is communicating about each subordinate's performance. In doing so, you should rely solely on your expertise and judgement to make the ratings. It is very important that each expert judge make his or her ratings totally independent of all other expert judges. The entire validity of this rating process rests on the fact that the expert judges have independently determined the appropriate ratings.

Getting Started

I have tried to make the process as simple as possible. If you just follow this package page by page you will complete all the steps necessary to make the appropriate ratings. First, check to see that you have the necessary materials.

1. A package titled "Situation Overview" that describes the setting of the simulation. It also includes an organizational chart for the organization created in this simulation. The organization chart includes pictures of the key subordinates whose performance you will evaluate.

2. A package titled "Behaviorally Anchored Rating Scales" (BARS). These are the performance scales that you will use to make your evaluations. They include the performance dimensions that are of interest and a detailed description of the types of behaviors that indicate a certain level of performance on each dimension.

3. A package titled "Evaluation Guide" which is a step by step guide to making your evaluations. It includes a background description of each scene and a rating matrix for you to make all your evaluations.

4. At least one videotape that is labeled with the scene numbers that are included on that tape. This will be either scenes 1-16, 17-32, 33-48 or 49-64.

5. A rater information form.

What Do You Do Next?

After insuring that you have the required materials, you should complete the following steps.

1. Review the "Situation Overview" to familiarize yourself with the context of the simulation.

2. Familiarize yourself with the performance dimensions by reviewing the BARS scales.

3. Find a quiet place where you have access to a television set and a VCR.

4. Read the background to the scene provided in the evaluation guide. Watch a scene carefully and then stop the tape and make the ratings following the steps provided. Note: Each scene was shot twice with the same characters responding differently to the same or a very similar situation. These scenes are referred to as scenes A and B. For example, there is a scene 1A and a scene 1B. Rate each

scene separately.

5. When you have completely rated a scene, restart the VCR and move to the next scene. Remember, we need you to rate each scene independently of all other scenes. You should not consider any information from a previous scene when rating a new scene. In other words each scene stands on its own and should be rated separately.

6. When you have completed the entire videotape, review each page to make sure you have completed all ratings.

7. After you have completed your ratings complete the Rater Information Form so that I can keep a record of the educational/experience background of the expert raters.

8. When you have completed steps 1-7 mail this package back to me at: **1030 NW 16th Street**
Gainesville, FL
32605

Thanks again for you help. I anticipate this project taking two hours or so of your time although it may take slightly longer for some. If at any time you have any questions, please don't hesitate to call me collect at **904-372-4933**.

SITUATION OVERVIEW

In this simulation, subjects are in the role of Leslie Wilder, newly appointed Chief of the Administrative and MIS Contracting Division for the Consolidated Federal Agency Contracting Office (CFACO). The CFACO is a fictitious organization that serves as the focal point for all the purchasing and supply activities that serve several major agencies of the federal government. In their position, they have a staff of individuals who manage the day to day activities of the procurement division. These individuals include:

Alice Garcia - Contract Compliance Manager
Carole Donaldson - Western Region Contracting Manager
Edward Montage - Central Region Contracting Manager
David Fredericks - Eastern Region Contracting Manager
Bill Jensen - Management Information Systems Contracting Manager

Each of these individuals manages a staff of 3 to 5 Contracting/Technical Specialists. Each also has a secretary/assistant. Within Leslie's division there are five major sections. The first three sections are the regional contracting sections divided by geographic region. The three contracting managers in these positions are responsible for all purchasing activities for regional offices throughout the United States. There is also a section that has responsibility for Management Information Systems procurement. Finally, there is a section whose sole responsibility is to insure compliance with government contracting directives.

During the simulation, subjects are introduced to their position as Leslie Wilder and then are given a packet of materials which represents administrative items in their in-basket. These materials are used to present managerial challenges (problem solving, administration, communication etc.) and to familiarize them more with the organization. All information in the in-baskets is background information that helps to set up the videotaped scenes the subjects view and that you are going to rate. These scenes are intended to portray the behavior of four subordinates, Alice, Bill, Carole and David. The fifth subordinate, Ed is in many scenes but his behavior is not of interest to us in this part of the simulation and so you do not need to rate him.

Each videotaped scene represents a behavioral episode for one or more subordinates. This is where you come in. We need our expert judges to watch each scene separately, determine which subordinates are behaving in a way that reflects on one or more of the key dimensions, and rate the level of performance on those dimensions using the BARS scales provided. Remember, each scene stands on its own and should be rated separately. There are several scenes where more than one of the key subordinates appear. Rate all subordinates whose actions in the scene allow you to judge their performance on one or more of the dimensions. The rating matrix for each scene is designed to allow you to rate any subordinate in the scene on any of the dimensions which you feel can be rated.

By way of summary let me emphasize several points.

1. Each scene stands on its own and should be rated separately. Don't let scenes that seem contradictory to previous scenes confuse you. The subjects will not see all the scenes you will see; they will see only a selected number and that is why each scene needs to be rated separately. Having a wide variety of scenes and behaviors to choose from provides us a great deal of flexibility in presenting subjects with different subordinate performance profiles.

2. Just because someone is in a scene doesn't mean their performance must necessarily be rated. You might see subordinates in a scene where their actions do not reflect on any of the performance dimensions.

3. Make sure you rate all subordinates in a scene if their behavior supports a rating on one or more performance dimensions.

4. You should only be concerned with rating the performance of Alice, Bill, Carole or David. Anyone else's performance is not the subject of this simulation. For example, you will see the subordinates interact with their subordinates. Your focus should be on how Alice, Bill, Carole or David respond in their roles as supervisors.

One last point about the scenes you will see. Sometimes the acting or camera work is not perfect. Don't let this disturb you, just rate what you see. When I actually use the videotapes in experiments subjects are so busy they usually don't notice these technical imperfections.

While this may seem confusing at first, I am sure that after you watch a few scenes it will become clearer. Remember, if you have any questions, just give me a call!

Thanks again for your help.

APPENDIX D
STUDY PROTOCOL

Day of The First Session

Note: As participants entered, they were assigned randomly to a group and the appropriate materials were distributed. Students who had not previously completed the demographic questionnaire did so at that time. When all participants were registered they were told:

> Thank you for participating in this research project. While this project may require more effort than other research projects you have participated in, I believe you will find the simulation both challenging and interesting. Those of you who complete the simulation will also find it rewarding in that your professor has been gracious enough to allow me to award participants up to 5 points of extra credit. Remember, you need to attend one session both this week and next and you will be required to complete a four page written assignment to receive any credit for participation. Your participation is voluntary and you may drop out at any time. However, only those who complete all requirements will receive any credit for participation. Check your schedules, make sure you can meet all requirements before we proceed. Are there any questions?

Please identify the location where you are sitting and place your name and participant number in the appropriate location. This information will be used to insure you receive proper course credit.

Please read the form titled "Introduction to The Simulation" while I read it aloud. **(Read form).** Are there any questions?

INTRODUCTION TO THE SIMULATION

This exercise will place you in the role of a manager in a large federal organization. While the name you will assume is not yours, it is important that you make decisions as if they were your own. This research is interested in both how individuals make decisions and how they respond to a wide variety of organizational situations. In each situation you confront, take the action and/or make the decision(s) you would make if you were actually confronted with this situation.

In your role as manager, you will report to a supervisor by the name of Art Bollinger. Throughout this simulation, the researcher will actually fill the position of Art Bollinger. As your supervisor, the researcher will not respond directly to any questions you have, however, any communication you would want to have with your supervisor should be documented in response to an item from the simulation. Also, as your supervisor the researcher will be compiling the results of your work and will be providing Dr Hall with the number of points you will receive for participation.

While most of what each participant sees in this simulation will be the same, there will occasionally be some differences. These differences may present certain participants situations which may be different from those of other participants. Please pay attention only to the materials presented to you and respond only to those materials. Responding to materials other than those presented to you could result in you receiving less credit for participation.

Information will be provided to you through both the contents of your in-basket, which is in the envelope that was handed to you when you entered the room, and by the television monitor you see in the front of the room. You should respond to information from both sources as you deem appropriate. It is important that you pay close attention to the actions of the individuals that you will observe on the television monitor. The colleagues and subordinates you will see will provide you important information for responding to the problems presented in this simulation.

This simulation has been designed so that you will be unable to complete all the work in your in-basket. As a result, it is important that you use your time wisely by prioritizing your actions based on the work that you feel is most important. While you have approximately one hour and 30 minutes to work on the in-basket you will occasionally be interrupted, as managers frequently are, and as a result you may find your priorities change throughout the scenario.

On the upper right hand corner of your in-basket envelope you will see a participant number printed on a small white label. This number will help us identify your work for administrative purposes. All materials for the simulation should be returned to this envelope. You will receive similar materials with the same number on it through out the exercise. Please do not put your name on any of these materials. We will keep track of you through your participant number. You can help us by recording your participant number and insuring that any materials you complete have the same number.

Note: The following form presents the accountability manipulation.

Please read the form titled "Requirements To Receive Course Credit". Different participants may be given different final assignments so they focus on different aspects of the simulation. If you have any questions, please ask me individually.

Please read the form titled "Consolidated Federal Agency Contracting Office In-Basket Exercise Overview" to yourself, while I read it aloud. (Pause) Are there any questions?

Note: The following form presents the rating outcome manipulation.

You are about to see a memo written from your supervisor Art Bollinger. This memo is representative of the types of materials you will see in the in-basket. This letter, like many of the other in-basket items, is very important for successful completion of the in-basket. One aspect of this simulation that is certainly realistic is that you are new to the organization and you really don't know a great deal about the way things are done in this organization. Please pay careful attention to materials like this since they provide information that will help you understand important aspects of the organization. Also, remember different students may receive different information and each student should respond only to the information provided them. If there are no questions, please read the memo.

Presentation of the simulation:

When you open your envelopes you will see all of the in-basket materials. While you should record your responses to the in-basket items directly on that item, you may find that there are responses you would like to take or decisions you would like to make that there is not appropriate spot to do so. Please use the white pad to record any in-basket responses, actions etc. you would like to record. You also have these yellow pads, please use these pages if you care to record items that specifically relate to the individual performance of one of your subordinates. **(Distribute Pads)** Are there any questions. Please wait for instructions from the video before you open your in-basket envelopes.

After 80 minutes:

Your time is up. Please gather all your materials together. Neatly tear off the white and yellow pages from the tablets provided. Make sure you put your participant number on the top right hand corner of each page. Place the white pages and all of your other materials in the envelope. Place the yellow pages on top and put a rubber band around the package and remain in your seat.

Day of The Second Session

Note: At the scheduled time participants were again welcomed and thanked for their participation. Each participant was handed the appropriate materials based on their participant number assigned during the first week's session. Participants were then told:

> Remember, this is the last session you will need to attend however you are still required to complete a short (approximately one-hour in preparation time) written assignment to receive any credit for participation. You can drop out at any time but they you will only receive credit for completing all requirements.
>
> Please identify the location where you are sitting and place your name and participant number in the appropriate location. This information will be used to insure you receive proper course credit.

Note: The following form presents the accountability manipulation.

> Please read the form titled "Requirements To Receive Course Credit". Different participants may be given different final assignments so they focus on different

aspects of the simulation. If you have any questions, please ask me individually.

Please read the form titled "Consolidated Federal Agency Contracting Office In-Basket Exercise Overview" to yourself, while I read it aloud. (Pause) Are there any questions?

Note: The following form presents the accountability manipulation.

Since approximately two weeks have elapsed since the last session, we have provided you excerpts from your notebook to familiarize you with any actions that occurred during that time period. Please read the form titled "Notebook Excerpts".

When you open your envelopes you will see all of the in-basket materials. While you should record your responses to the in-basket items directly on that item, you may find that there are responses you would like to take or decisions you would like to make that there is not appropriate spot to do so. Please use the white pad to record any in-basket responses, actions etc. you would like to record. You also have these yellow pads, please use these pages if you care to record items that specifically relate to the individual performance of one of your subordinates. (Distribute Pads) Are there any questions.

After 80 Minutes:

Your time is up. Please gather all your materials together. Neatly tear off the white and yellow pages from the tablets provided. Make sure you put your participant number on the top right hand corner of each page. Place the white pages and all of your other materials in the envelope. Place the yellow pages on top and put a rubber band around the package and remain in your seat.

Gathering performance ratings:

Your next task is to rate the performance of each subordinate on a series of performance dimensions. As we suggested at the beginning of this simulation, we would like you to evaluate their performance over both sessions of this simulation. Performance evaluation is a critical aspect of a manager's job and as such you should take care to perform these evaluation as carefully and accurately as possible. Since some of you will need a record of your performance ratings to assist you in your written assignment, you are being provided a sheet to record your scores on.

After completion of the evaluations, participants were given the questionnaire containing the manipulation checks, the observational accuracy instrument, and the "Reactions To The Simulation" questionnaire. Finally, participants were told:

> Thank you for your participation The purpose of the study was to examine how different organizational situations affected how each of you reacted to organizational problems and evaluated individual performance. You need to pick up an information sheet that describes how to complete your written assignment. These assignments are due to BUS 219 by 12:00 P.M. on October 1.

APPENDIX E
ACCOUNTABILITY MANIPULATION

Requirement to Receive Course Credit
(Accountable Condition)

In addition to attending both sessions of the simulation and completing the related in-baskets, you will be required to complete a performance evaluation on each of your subordinates during the second session. In this evaluation, you will be rating each of your key subordinates on four performance dimensions:

1. Administrative abilities - How well does the subordinate write, attend to detail in their administrative products, and meet timelines for work assignments.

2. Work Effort - How well or often the subordinate volunteers for challenging work assignments, sacrifices personal convenience to accomplish tasks, handles tasks beyond their normal area of responsibility, or handles crises.

3. Supervision - How well does the subordinate motivate, develop, and inform their subordinates.

4. Coordination and negotiation - How well does the subordinate maintain good working relationships, handle conflict, deal with stressful interpersonal situations, or demonstrate receptivity to inputs of others.

You will also be required to complete a written assignment where you will provide written justification for the performance ratings you assign. This justification is an important part of determining the amount of extra credit you will receive. The purpose of this assignment is to justify your evaluation of the performance of each subordinate on each performance dimensions. You will be evaluated on your justification within the context of the material presented within the simulation. If your ratings are not understood by researchers, your written justifications will be used to understand your rationale for them. If additional clarification is required, you will be asked to meet with researchers to provide further justification. More information on format is provided on the following page.

FINAL ASSIGNMENT

To receive credit for participation in this simulation, you must provide written justification for the performance ratings you assigned. This justification will be an important part of determining the course credit you will receive for this assignment. The purpose of this is to justify your evaluation of the performance of each subordinate on each of the performance dimensions.

Here are the instructions to follow when preparing your assignment:

1. Justify your ratings for each subordinate on a separate page.

2. Answers should be typed or legibly handwritten.

3. Include your name and participant number on each page.

4. **Responses must be dropped off at BUS 219 by 12:00 P.M. Friday, 1 October.**

5. Use the following format

   - Subordinate Name:

   - Performance Dimension:

   - Rating Assigned:

   - Justification: Provide justification for the ratings you assigned including specific examples of subordinate performance that supports that rating.

   - Repeat this format until you have discussed your ratings on all four performance dimensions. Then proceed to the next subordinate until you have justified all of your ratings.

Requirement to Receive Course Credit
(Non-accountable condition)

In addition to attending both sessions of the simulation and completing the related in-baskets, you will be required to complete a performance evaluation on each of your subordinates during the second session. In this evaluation, you will be rating each of your key subordinates on four performance dimensions:

1. Administrative abilities - How well does the subordinate write, attend to detail in their administrative products, and meet timelines for work assignments.

2. Work Effort - How well or often the subordinate volunteers for challenging work assignments, sacrifices personal convenience to accomplish tasks, handles tasks beyond their normal area of responsibility, or handles crises.

3. Supervision - How well does the subordinate motivate, develop, and inform their subordinates.

4. Coordination and negotiation - How well does the subordinate maintain good working relationships, handle conflict, deal with stressful interpersonal situations, or demonstrate receptivity to inputs of others.

The performance ratings you assign will remain anonymous and will be combined with the ratings of other participants to provide average ratings that will be considered for the purposes of this research. While you should take care in making your ratings, you will not have to justify your ratings to anyone.

After completion of the simulation you will be given an assignment to complete out of class. More information on this assignment is provided on the following page.

FINAL ASSIGNMENT

To receive credit for participation in this simulation, you must respond to a series of questions concerning your reactions to the simulation. Please take your time in responding to this questionnaire carefully and thoughtfully. Your inputs are important in developing these type of simulations for classroom use.

Here are the instructions to follow when preparing your assignment:

1. Respond to each question on a separate page.

2. Answers should be typed or legibly handwritten.

3. Include your name and participant number on each page.

4. **Responses must be dropped off at BUS 219 by 12:00 P.M. Friday, 1 October.**

1. Did you feel that the situations presented were realistic and reflected the types of situations managers deal with on a daily basis? In your response, please address the following issues:
    - Were organizational members presented in a realistic manner?
    - Were the types of problems realistic?
    - Were decisions you made reflective of decisions you would make if you were actually faced with such a situation?
    - If not, how were they different?

2. Did the simulation provide adequate information to help you make decisions? In your response, please address the following issues:
    - Was it clear to you when a decision was required?
    - Were you able to piece several sources of information to determine an appropriate response?

3. If a simulation like this were used as a regular part of a course in management, would it be helpful in your understanding and application of course materials? Please thoroughly explain your answer.

4. Discuss the strengths and weaknesses of the simulation. In your response, please address the following issues:
    - What did you like best about this simulation?
    - What did you like least?
    - How could the simulation be improved to make it more interesting and a better educational experience?

APPENDIX F
RATING OUTCOME MANIPULATION

CONSOLIDATED FEDERAL AGENCY
CONTRACTING OFFICE

Memo

(Presented to Subjects in the Inflationary Condition)

To: Leslie Wilder

From: Art Bollinger

Date: May 13

Welcome to the organization. I know you must be feeling a little intimidated by the stack of work you have in front of you but I have every confidence that you will do a first rate job. Before you begin however, there is one thing I want to make you aware of as you begin your new supervisory duties.

As you know, performance evaluation is but one of your many duties. This area has however been a problem for us in the past. In the last 5 years, not one of the department managers in your division has been promoted to Division Chief Level. I have looked into this matter and I believe this may be due to the rating policies of your predecessor, Bob Ward. The ratings of personnel in your division have consistently been the lowest in our organization. I have often been impressed with the work of members of your staff and so those ratings surprise me. I have compiled some results of those ratings from the last evaluation period to give you an example of what I mean.

CFACO IS AN EQUAL OPPORTUNITY EMPLOYER

CONSOLIDATED FEDERAL AGENCY CONTRACTING OFFICE

Memo

DIMENSIONS	DIVISION MEAN	CFACO MEAN
Administration	4.8	6.2
Work Effort	3.9	5.9
Supervision	4.5	6.1
Coordination & Negotiation	4.2	6.2

I am not sure of the reason for the lower ratings, but I do know it has made it difficult for us to hire someone from your division for positions of advanced responsibility. The next performance evaluation will occur in a few weeks, so you will get a chance to make your own ratings based on the performance you observe.

I know as you work through your in-basket, and meet your staff, you will have many challenging problems to tackle. I just wanted to make you aware of a situation that is of great concern to me, as I am sure it will be to you. Good luck!

Art

CONSOLIDATED FEDERAL AGENCY
CONTRACTING OFFICE

Memo

(Presented to Subjects in the Equitable Treatment Condition)

To: Leslie Wilder

From: Art Bollinger

Date: May 13

Welcome to the organization. I know you must be feeling a little intimidated by the stack of work you have in front of you but I have every confidence that you will do a first rate job. Before you begin however, there is one thing I want to make you aware of as you begin your new supervisory duties.

As you know, both public and private employers have come under intense scrutiny in how their personnel practices affect minorities and women. CFACO has not escaped this attention, especially recently when a suit was filed by a member of another division claiming that women are being unfairly treated because they receive on average significantly lower performance evaluations than men. I have reviewed the evaluations of women in our organization and it is clear that they have consistently been rated lower. Since the performance evaluations are used for promotion decisions, this has been a factor in why fewer women are promoted. While performance evaluation is only one of your many duties it is one that I consider very important. The next performance evaluation will occur in a few weeks, so you will get a chance to make your own ratings based on the performance you observe.

I know as you work through your in-basket, and meet your staff, you will have many challenging problems to tackle. I just wanted to make you aware of a situation that is of great concern to me, as I am sure it will be to you. Good luck!

Art

CONSOLIDATED FEDERAL AGENCY
CONTRACTING OFFICE

Memo

(Presented to Subjects in the Accuracy Condition)

To: Leslie Wilder

From: Art Bollinger

Date: May 13

Welcome to the organization. I know you must be feeling a little intimidated by the stack of work you have in front of you but I have every confidence that you will do a first rate job. Before you begin however, there is one thing I want to make you aware of as you begin your new supervisory duties.

As you know, performance evaluation is but one of your many duties. This area has however been a problem for us in the past. In the last 5 years it seems that all performance ratings in the organization have become inflated to the point where everyone gets the top rating on all the different performance dimensions. As a result this makes it very difficult for us to make important personnel decisions using these evaluations. For example, when I reviewed the most recent performance evaluations of your staff I can't tell any difference in their performance levels. I have to tell you if we have an opportunity to promote someone, we couldn't use the evaluations to help make the decision. Additionally, the most recent evaluations don't give me any idea what the strengths and weaknesses are for each individual. Since the ratings for each different performance dimension are the same, we can't use them to make any decisions. Just the other day I wanted to put together a special project team and I was looking for someone with strong administrative skills, the performance evaluations did not provide me the necessary information to make a decision. The next performance evaluation will occur in a few weeks, so you will get a chance to make your own ratings based on the performance you observe.

I know as you work through your in-basket, and meet your staff, you will have many challenging problems to tackle. I just wanted to make you are aware of a situation that is of great concern to me, as I am sure it will be to you. Good luck!

Art

CONSOLIDATED FEDERAL AGENCY
CONTRACTING OFFICE

Memo

(**Presented to Subjects in the Comparison Group**)

To: Leslie Wilder

From: Art Bollinger

Date: May 13

Welcome to the organization. I know you must be feeling a little intimidated by the stack of work you have in front of you but I have every confidence that you will do a first rate job. Before you begin however, there are some important issues in the organization that I want to make you aware of before you begin your new duties.

There are three challenges I see this organization facing. First, our budget has been cut in three of the last four years. This means we are asked to meet greater procurement needs with fewer resources. This requires each of our division chiefs to concentrate on efficiency in the performance of their duties.

Second, our contractor selection process is coming under increasing pressure by congress to reduce our reliance on foreign manufacturers of many goods. This is extremely difficult in light of the fact that the number of US suppliers seems to be constantly diminishing. It is

CFACO IS AN EQUAL OPPORTUNITY EMPLOYER

imperative that we keep this situation in mind as we approach our duties throughout the next year.

Third, the rapid turnover of our contracting personnel is becoming a serious problems. I want each Division Chief to focus on ways for us to better understand why so many contracting specialists are leaving and how we can improve our retention of these important human resources. You will be hearing more from me on this issue over the next several days and weeks.

I know as you work through your in-basket, and meet your staff, you will have many challenging problems to tackle. I just wanted to make you aware of some situations that are of great concern to me, as I am sure it will be to you. Good luck!

Art

APPENDIX G
RATING FORMAT

RATING FOR ALICE GARCIA

Please follow these steps for rating the performance of Alice Garcia:

First, read the definition of the Administrative dimension on the rating scales provided.

Second, read the description of behaviors for the three ranges of performance.

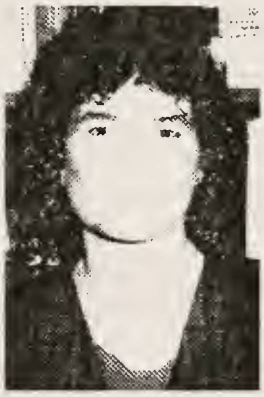
Third, decide which of the three descriptions (HIGH, MODERATE, or LOW) best fits Alice's Administrative performance.

Fourth, Determine which ONE NUMBER within that range best represents Alice's Administrative performance.

Fifth, write that NUMBER below under the rating column on the line to the right of the Administrative dimension.

When you complete the rating for the Administrative dimension, follow the same procedures to rate the Coordination and Negotiation, Supervision, and Work Effort dimensions.

Dimension	Rating (1-7)
Administrative	______
Coordination and Negotiation	______
Supervisory	______
Work Effort	______



When you have completely rated the performance of Alice, turn to the next page.

RATING FOR BILL JENSEN

Please follow these steps for rating the performance of Bill Jensen:

First, read the definition of the Administrative dimension on the rating scales provided.

Second, read the description of behaviors for the three ranges of performance.

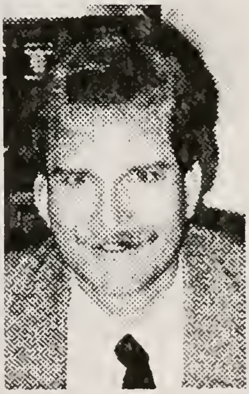
Third, decide which of the three descriptions (HIGH, MODERATE, or LOW) best fits Bill's Administrative performance.

Fourth, Determine which ONE NUMBER within that range best represents Bill's Administrative performance.

Fifth, write that NUMBER below under the rating column on the line to the right of the Administrative dimension.

When you complete the rating for the Administrative Dimension, follow the same procedures to rate the Coordination and Negotiation, Supervisory, and Work Effort Dimensions.

Dimension	Rating (1-7)
Administrative	______
Coordination and Negotiation	______
Supervisory	______
Work Effort	______



When you have completely rated the performance of Bill, turn to the next page.

RATING FOR CAROLE DONALDSON

Please follow these steps for rating the performance of Carole Donaldson:

First, read the definition of the Administrative dimension on the rating scales provided.

Second, read the description of behaviors for the three ranges of performance.

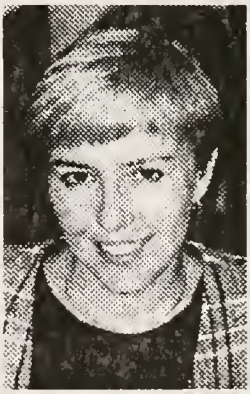
Third, decide which of the three descriptions (HIGH, MODERATE, or LOW) best fits Carole's Administrative performance.

Fourth, Determine which ONE NUMBER within that range best represents Carole's Administrative performance.

Fifth, write that NUMBER below under the rating column on the line to the right of the Administrative dimension.

When you complete the rating for the Administrative Dimension, follow the same procedures to rate the Coordination and Negotiation, Supervisory, and Work Effort Dimensions.

Dimension	Rating (1-7)
Administrative	______
Coordination and Negotiation	______
Supervisory	______
Work Effort	______



When you have completely rated the performance of Carole, turn to the next page.

RATING FOR DAVID FREDERICKS

Please follow these steps for rating the performance of David Fredericks:

First, read the definition of the Administrative dimension on the rating scales provided.

Second, read the description of behaviors for the three ranges of performance.

Third, decide which of the three descriptions (HIGH, MODERATE, or LOW) best fits David's Administrative performance.

Fourth, Determine which ONE NUMBER within that range best represents David's Administrative performance.

Fifth, write that NUMBER below on the line under the rating column to the right of the Administrative dimension.

When you complete the rating for the Administrative Dimension, follow the same procedures to rate the Coordination and Negotiation, Supervisory, and Work Effort Dimensions.

Dimension	Rating (1-7)
Administrative	______
Coordination and Negotiation	______
Supervisory	______
Work Effort	______

When you have completely rated the performance of David, check to insure you have rated all four subordinates on all four dimensions. Make sure you have provided a rating of between 1 and 7 to indicate your evaluation of the performance of each subordinate on all dimensions. When you have completed this evaluation, remain in your seat and wait for the next task.

APPENDIX H
DEMOGRAPHIC QUESTIONNAIRE

NAME: ______________________________

Note: Names will be used for administrative purposes only and will be removed at the conclusion of the experiment.

CFACO IN-BASKET EXERCISE PARTICIPANT WORK SHEET

PARTICIPANT NUMBER _____

INSTRUCTIONS: Please fill out this questionnaire as completely and as accurately as possible. Information provided on this questionnaire will be used for research purposes only. Please note the participant number listed above, we will use this number to identify you throughout this project. The researchers will have a record of this number should you forget it.

AGE: _____

SEX: Male_____ Female_____

NUMBER OF YEARS OF (TOTAL) WORK EXPERIENCE: _____

NUMBER OF YEARS OF MANAGEMENT EXPERIENCE: _____
(Include work experience where you were directly responsible for managing people, budgets, projects etc.)

NUMBER OF YEARS OF SUPERVISORY EXPERIENCE: _____

UNDERGRADUATE MAJOR: ______________________

PREVIOUS GRADUATE EDUCATION: ______________________

APPENDIX I
MANIPULATION CHECK

The next series of questions asks you to indicate whether you remember seeing specific memos and letters in your in-basket. Remember, since the in-basket included different items for different participants you may have seen a memo that the person sitting next to you did not see. Circle YES if you remember seeing the information referred to, circle NO if you do not remember seeing that information.

I remember seeing the following memos or letters.......

a) a memo from Art Bollinger comparing previous performance evaluations of personnel within the Administrative & MIS Division with those of personnel in other divisions.	YES	NO
b) a memo from Art Bollinger discussing the major organizational challenges to CFACO.	YES	NO
c) a memo from Art Bollinger discussing the ratings received by women within CFACO.	YES	NO
d) a memo from Art Bollinger discussing the usefulness of previous performance appraisals in making important organizational decisions.	YES	NO
e) a letter from a woman who used to work within the division and who was concerned about the impact that previous performance ratings she had received were having on a job she was trying to get.	YES	NO
f) a letter from Wanda Lane which compared the average performance ratings of personnel from each division within CFACO.	YES	NO
g) a letter from the HR Director at another federal organization that discussed the results of a recent selection board where they considered three members of CFACO for a position within their organization.	YES	NO

APPENDIX J
OBSERVATIONAL ACCURACY MEASURE

INSTRUCTIONS FOR COMPLETING THESE FORMS

Your next task is to decide which of the following scenes you remember observing over the last two weeks.

You will find below a list of 32 statements that describe events that might or might not have occurred at some point during one of simulation sessions. Decide whether these events actually took place and put a "Yes" in front of each scene you saw and a "No" in front of each scene you did not see.

_____ 1. Alice Garcia worked over the weekend to complete a review of installation schedules for Bill Jensen.

_____ 2. Carole Donaldson decided not to take vacation time so she could stay at CFACO and help a colleague in another division during a meeting with a contractor.

_____ 3. David Fredericks asked that his name not be considered for project officer for the CFACO conference for key federal agency contacts.

_____ 4. David Fredericks was over a week late in turning in the previous month's trip reports.

_____ 5. Bill Jensen was unwilling to assist David Fredericks in resolving a dispute between a contractor and a federal client.

_____ 6. Mike Jones told Leslie that Alice Garcia had done a great deal to develop his confidence and encouraged him to apply for the White House Fellows program.

_____ 7. Alice Garcia made sure her subordinates were well prepared to discuss important CFACO issues in a meeting with Art Bollinger.

_____ 8. David Fredericks maintained his composure despite being falsely accused by a member of the finance division of having the wrong funding codes on a contract.

_____ 9. Carole Donaldson took positive steps to address concerns of a federal client in Los Angeles who had written to Art Bollinger about past problems she has had in dealings with CFACO.

_____ 10. David Fredericks was unwilling to fill in during a contract supplier selection in the services division.

_____ 11. Bill Jensen positively considered Carole Donaldson's advice about dealing with a customer that had called Carole to complain about support they were receiving for their MIS system.

_____ 12. During a staff meeting, Alice Garcia accused her colleagues of going behind her back about their concerns that her section was disrupting their work flow.

_____ 13. Bill Jensen rejected Ed Montage's advice about a relationship with a contractor that had become strained because Bill had encouraged other contractors to bid on a project.

_____ 14. David Fredericks was supportive of a subordinate who was struggling to get his work done despite having difficulties at home.

_____ 15. Carole Donaldson worked over the weekend to complete a review of installation schedules for Bill Jensen.

_____ 16. Carole Donaldson was unaware that one of her subordinates was considered the "resident expert" for CFACO reporting procedures.

_____ 17. Carole Donaldson rejected Alice Garcia's attempt to discuss Carole's comments on the cost of minority contracts.

_____ 18. Bill Jensen was unwilling to help Wendy deal with a frustrated client who had been trying to reach Ed Montage's section.

_____ 19. David Fredericks missed a 10:00 A.M. appointment as a result of taking care of some personal errands.

_____ 20. Bill Jensen rearranged his subordinate's travel schedule in an attempt to reduce the amount by which his section exceeded the travel budget.

_____ 21. Alice Garcia approached Carole Donaldson to discuss Carole's feelings about Alice holding further discussions with a small business supplier who had unsuccessfully bid on a contract.

_____ 22. Bill Jensen wrote a letter on the MIS system to senior personnel in CFACO's client organization that drew angry reaction from some subordinates in those organizations.

_____ 23. Alice Garcia was not ready at the staff meeting to report on recent changes in federal legislation.

_____ 24. Carole Donaldson was critical of a letter written by a subordinate and decided to write the letter herself.

_____ 25. Carole Donaldson was late turning in her monthly status sheet to Wendy.

_____ 26. Carole Donaldson suggested to Wanda Lane that if she wasn't allowed to make an emergency hire, she would send her work to Wanda's division.

_____ 27. David Fredericks aggravated a dispute between a client who claimed the equipment they were sent didn't work and a contractor who felt the client was improperly using the equipment.

_____ 28. Alice Garcia praised a subordinate for showing teamwork in the written comments he had submitted on a contract prepared by the Eastern Region.

_____ 29. Bill Jensen resolved a dispute between Ed Montage and a member of the finance division by asking both individuals to put themselves in the position of the other.

_____ 30. David Fredericks invited Leslie to attend a luncheon where he was going to recognize the efforts of his staff.

_____ 31. Alice Garcia volunteered to work on improving the CFACO management development program.

_____ 32. Bill Jensen gave work on a contract that his subordinate was unable to finish to a different subordinate.

APPENDIX K
POST-EXPERIMENT QUESTIONNAIRE

Your Reactions to the In-basket Exercise

Show how much you agree or disagree with each statement below by circling a number from 1 to 5 at the right.

1=strongly disagree
2=disagree
3=neither agree nor disagree
4=agree
5=strongly agree

a) The in-basket materials were clear and understandable.	1	2	3	4	5
b) There was enough time to resolve the most important problems in the in-basket.	1	2	3	4	5
c) I understand how the videotaped events are related to the in-basket materials.	1	2	3	4	5
d) The videotaped interruptions interfered with my concentration on the in-basket materials.	1	2	3	4	5
e) The videotaped events presented useful information for resolving problems in the in-basket.	1	2	3	4	5
f) I was bored by this simulation.	1	2	3	4	5
g) I was stimulated and challenged by this simulation.	1	2	3	4	5
h) This was an interesting and realistic exercise.	1	2	3	4	5
i) I would have done things differently if this were a real work situation.	1	2	3	4	5
j) As I worked through the simulation, I was aware of the type of information that I would need to help me complete the written assignment	1	2	3	4	5
k) All in all, this exercise is useful for illustrating the demands of managerial work.	1	2	3	4	5

Show how much you agree or disagree with each statement below by circling a number from 1 to 5 at the right.

1=strongly disagree
2=disagree
3=neither agree nor disagree
4=agree
5=strongly agree

---

l. I found myself thinking about different aspects of the simulation between the first and second sessions. 1 2 3 4 5

m. I was looking forward to the second session. 1 2 3 4 5

n. When making decisions in the simulation, I frequently considered and debated the merits of alternative responses to particular problems. 1 2 3 4 5

o. I am concerned that my responses to the in-basket or on the paper due next week may lead to me not receiving all 5 extra credit points for participation in this simulation. 1 2 3 4 5

p. I found the inability to immediately communicate with colleagues and subordinates in the simulation very frustrating. 1 2 3 4 5

q. Between sessions, I discussed some of the situations I saw in the first part of the simulation with someone who was not participating in this project. 1 2 3 4 5

When responding to the next set of questions, consider _all_ of the information you have seen in this simulation by trying to recall the number of positive and negative pieces of information you saw on each subordinate. In front of each question place a number that best represents your response.

_____ r. How many times do you remember Alice doing something that positively reflected on her overall performance.

_____ s. How many times do you remember Alice doing something that negatively reflected on her overall performance.

_____ t. How many times do you remember Bill doing something that positively reflected on his overall performance.

_____ u. How many times do you remember Bill doing something that negatively reflected on his overall performance.

_____ v. How many times do you remember Carole doing something that positively reflected on her overall performance.

_____ w. How many times do you remember Carole doing something that negatively reflected on her overall performance.

_____ x. How many times do you remember David doing something that positively reflected on his overall performance.

_____ y. How many times do you remember David doing something that negatively reflected on his overall performance.

APPENDIX L
ATTENTIVENESS RATING SCALE

Attentiveness Rating Scale

Instructions:

Please rate each subject on the attentiveness scale provided below. Provide rating one for each subject based on a scale from 1 to 3 with 1 representing low attentiveness and 3 representing high attentiveness.

Attentiveness is the level of alertness and interest displayed by the subject when viewing performance events as indicated by head position, expression, posture, and note taking reaction.

High Attentiveness

Alertly looks towards TV screen throughout the performance episode (except when taking notes) and responds to performances through verbal or non-verbal expressions

Appears interested in the action taking place on the screen by sitting upright and stopping work on non-related in-basket materials

Intently considers performance information by appearing thoughtful after performance episode

Takes notes during or after the performance episode on the form provided

Low Attentiveness

Rarely looks at the TV screen throughout the performance episode and appears oblivious to specifics aspects of performance

Appears disinterested in the action taking place on the screen by slouching in their seat and/or continuing to work on other in-basket items

Appears unaffected by performance information and continues to work on other material without pause

Does not appear to take any notes about performance information

APPENDIX M
NOTE TAKING RATING SCALE

Note Taking Comprehensiveness

Instructions:

Please rate each subject on the note taking comprehensiveness scale provided below. Provide one rating for each subject based on a scale from 0 to 4 with 0 representing low note taking comprehensiveness and 4 representing high note taking comprehensiveness.

Note Taking Comprehensiveness is indicated by the quantity and quality of notes taken by the subject on performance related information

High Note Taking Comprehensiveness

Took notes on most performance episodes

Notes were of sufficient quality to be used in performance assessment by including information as to who was involved, a description of the behavior, and pertinent events surrounding the episode.

Notes were organized in a logical manner such as by subordinate or by dimension

Low Note Taking Comprehensiveness

Took only a few notes on performance episodes

Notes were not of sufficient quality to support use in performance assessment because they failed to include information such as who was involved, a description of the behavior, or pertinent events surrounding the episode

Notes were not organized in a logical manner

BIOGRAPHICAL SKETCH

Neal Philip Mero was born August 19, 1955, in Hinsdale, Illinois. He is a career Air Force member who enlisted in the Air Force in 1975, was commissioned in 1981, and currently holds the rank of Major. He received his B.S. in business and management from the University of Maryland in 1981 and an M.B.A. from the University of Montana in 1985. He has held a variety of positions in the Air Force including a position as a faculty member in the Department of Management at the United States Air Force Academy. He is married and he and his wife Cheryl have two children, Samantha and Brent.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Stephan J. Motowidlo, Chairman
Associate Professor of Management

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy

Henry Tosi

Henry J. Tosi
Professor of Management

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy

Jerald W. Young
Associate Professor of Management

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy

Barry R. Schlenker
Professor of Psychology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy

Linda M Crocker

Linda M. Crocker
Professor of Foundations of
Education

This dissertation was submitted to the Graduate Faculty of the Department of Management in the College of Business Administration and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

April, 1994

Dean, Graduate School

LD
1780
1994
.M567

UNIVERSITY OF FLORIDA
3 1262 08553 8667